N° d'enregistrement
au C.N.R.S.

THESE

DE

DOCTORAT D'ETAT ES-SCIENCES MATHEMATIQUES

présentée

A L'UNIVERSITÉ PARIS VI

par

Raymond TRÉMOLIÈRES

pour obtenir

le grade de Docteur Es-Sciences

1ère Thèse

INEQUATIONS VARIATIONNELLES : EXISTENCE, APPROXIMATION, RÉSOLUTION

2ème Thèse

Propositions données par la Faculté

Soutenues le 27 octobre 1972 devant la Commission d'examen

MM. LIONS *Président*

HERVÉ, HAUGAZEAU } *Examinateurs*

N° d'enregistrement
au C.N.R.S.

THESE

DE

DOCTORAT D'ETAT ES-SCIENCES MATHEMATIQUES

présentée

A L'UNIVERSITÉ PARIS VI

par

Raymond TRÉMOLIÈRES

pour obtenir

le grade de Docteur Es-Sciences

1ère Thèse

INEQUATIONS VARIATIONNELLES : EXISTENCE, APPROXIMATION, RÉSOLUTION

2ème Thèse

Propositions données par la Faculté

Soutenues le 27 octobre 1972 devant la Commission d'examen

MM. LIONS *Président*

HERVÉ, HAUGAZEAU } *Examinateurs*

UNIV.
LEIDEN
BIBL.

Haec dies quam fecit Dominus...

Ps. 118 - 24

INTRODUCTION

Ce travail est consacré à l'existence, l'approximation et la résolution de trois types d'inéquations:

- inéquations stationnaires,
- inéquations paraboliques,
- inéquations hyperboliques ou bien posées au sens de Petrowski.

Dans le Chapitre I, nous montrons comment l'on peut établir l'existence de solutions pour une large catégorie d'inéquations variationnelles elliptiques en utilisant le Théorème de minimax de Ky-Fan-Sion : ceci repose sur la considération que le problème de trouver $u \in K$, convexe fermé d'un espace de Hilbert V tel que :

$$(I) \qquad (Au, v-u) + j(v) - j(u) \geqslant (f, v-u) \ , \quad \forall\, v \in K$$

- où A est un opérateur linéaire, monotone de V dans V' dual de V,
  j(v), une fonctionnelle semi-continue inférieurement et convexe,
  f , un élément de V' et (,) est le produit scalaire de la dualité entre V et V' -

est équivalent à trouver un point-selle de

$$L(u,v) = (Au, v-u) + j(v) - j(u) - (f, v-u) \qquad \text{sur } K \times K.$$

Dans le Chapitre II, nous étudions l'approximation de l'inéquation (I) par différences finies. On donne un certain nombre d'hypothèses générales qui permettent de contrôler dans quelle mesure un problème "approché" de (I) est correctement posé. On étudie en particulier l'approximation de quelques convexes fréquemment rencontrés dans les modèles de la physique des milieux continus.

Au Chapitre III, nous étudions l'existence pour les problèmes paraboliques. Nous montrons que l'utilisation du Théorème de Ky-Fian-Sion permet en particulier d'établir l'existence de solutions faibles.

Le Chapitre IV est consacré à l'approximation des inéquations paraboliques par différences finies et par divers schémas.

Dans le Chapitre V, nous étudions l'existence pour les inéquations hyperboliques ou bien posées au sens de Petrowski et dans le Chapitre VI, nous examinons l'approximation de ces problèmes par différences finies.

Au Chapitre VII, qui termine ce travail, nous donnons de nombreux exemples numériques de résolution des divers problèmes étudiés dans les chapitres précédents. Dans ce but, un grand nombre de méthodes traditionnelles ont été mises en oeuvre et quelques méthodes ou schémas originaux ont été introduits. Afin de clarifier la présentation, nous avons reporté à la fin de ce Chapitre, une conclusion générale concernant la valeur des méthodes utilisées.

--- § ---

Le sujet de ce travail me fût proposé par Monsieur LIONS que je remercie pour ses conseils, ses nombreux encouragements, et son soutien amical. Qu'il veuille bien trouver ici l'expression de ma profonde gratitude.

Je remercie Monsieur HAUGAZEAU d'avoir bien voulu examiner ce travail et Monsieur HERVE de m'avoir proposé un second sujet.

Que tous mes amis de l'I.R.I.A. soient ici remerciés pour leur très précieuse et très amicale collaboration.

La dactylographie a été réalisée par Mademoiselle Maryse GRANGEARD que je remercie très sincèrement.

L I S T E D E S C H A P I T R E S

CHAPITRE I : Existence et unicité pour les problèmes elliptiques

CHAPITRE II : Approximation des problèmes elliptiques

CHAPITRE III : Existence et unicité pour les problèmes paraboliques

CHAPITRE IV : Approximation des problèmes paraboliques

CHAPITRE V : Existence et unicité pour les problèmes hyperboliques ou bien posés au sens de Petrowski

CHAPITRE VI : Approximation des problèmes hyperboliques ou bien posés au sens de Petrowski

CHAPITRE VII : Résolution numérique de quelques problèmes

---- § ----

PLAN DETAILLE

		Pages
I.	EXISTENCE ET UNICITE POUR LES PROBLEMES ELLIPTIQUES	1
	1. Formulation abstraite	1
	2. Résultats préliminaires	3
	3. Existence	6
II.	APPROXIMATION DES PROBLEMES ELLIPTIQUES	8
	1. Théorie	8
	1.1. Cadre fonctionnel	8
	1.2. Approximation de K (ou de V lorsque K = V)	9
	1.3. Approximation de la forme $a(u,v)$	9
	1.4. Approximation de $j(v)$	10
	1.5. Approximation de f	12
	1.6. Problèmes approchés	12
	1.7. Convergence faible	13
	1.8. Convergence forte	14
	1.9. Définition fréquente du problème discrétisé	15
	2. Approximation externe	16
	2.1. Discrétisation de Ω	16
	2.2. Discrétisation de V, l'espace V_h et le produit scalaire $((,))_h$	18
	2.3. L'opérateur de prolongement p_h et la consistance de $\{p_h, V_h, V\}$	18
	2.4. L'opérateur r_h et la convergence de $\{p_h, V_h, V\}$	19
	2.5. Discrétisation de quelques convexes	19
	2.6. Discrétisation de $a(u,v)$	28
	2.7. Discrétisation de $j(v)$	28
	2.8. Discrétisation de f	30
III.	EXISTENCE ET UNICITE POUR LES PROBLEMES PARABOLIQUES	32
	1. Formulation abstraite	32
	2. Résultats préliminaires	33
	3. Existence dans le cas borné : solutions fortes	36
	4. Existence dans le cas borné : solutions faibles	37
	5. Problème pseudo-instantanés équivalents	38
	6. Problèmes instantanés équivalents	39
	7. Unicité	41

IV. APPROXIMATION DES PROBLEMES PARABOLIQUES 42

1. Théorie 42

1.1. Cadre fonctionnel 43
1.2. Approximation de $\mathcal{K}$ (ou de $W_o(0,T)$ lorsque $\mathcal{K} = W_o(0,T)$) 44
1.3. Approximation de $a(t;u,v)$ 45
1.4. Approximation de $j(t,v)$ 46
1.5. Approximation de $f(t)$ 47
1.6. Schémas de discrétisation 47
1.7. Lemmes préliminaires 49
1.8. Existence et stabilité pour le schéma I 51
1.9. Convergence faible du schéma I 52
1.10. Convergence forte du schéma I 53
1.11. Existence et stabilité pour le schéma II 54
1.12. Convergence faible du schéma II 56
1.13. Convergence forte du schéma II 57
1.11. Existence et stabilité pour le schéma III 59
1.12. Convergence faible du schéma III 66
1.13. Convergence forte du schéma III 69

2. Vérification des hypothèses 72

2.1. Hypothèses sur $\mathcal{K}$ 72
2.2. Hypothèses sur $a(t;u,v)$ 77
2.3. Hypothèses sur $j(t,v)$ 79
2.4. Hypothèses sur f 80

V. EXISTENCE ET UNICITE POUR LES PROBLEMES HYPERBOLIQUES OU BIEN POSES AU SENS DE PETROWSKI 81

1. Formulation abstraite 81
2. Résultats préliminaires 82
3. Existence dans le cas borné : solutions fortes 84
4. Existence dans le cas borné : solutions faibles 85
5. Problèmes pseudo-instantanés 86
6. Problèmes instantanés 86
7. Unicité 87

VI. APPROXIMATION DES PROBLEMES HYPERBOLIQUES OU BIEN POSES AU SENS DE PETROWSKI 88

1. Théorie 88

1.1. Cadre fonctionnel 89
1.2. Approximation de $\mathcal{K}$ (ou de $Y_o(0,T)$ lorsque $\mathcal{K} = Y_o(0,T)$) 89

1.3. Approximation de $a(t;u,v)$ 90
1.4. Approximation de $j(t,v)$ 91
1.5. Approximation de $f(t)$ 91
1.6. Schémas de discrétisation 91
1.7. Existence et stabilité pour le schéma I 95
1.8. Convergence faible du schéma I 98
1.9. Existence et stabilité pour le schéma II 101
1.10. Convergence faible du schéma II 104
1.11. Convergence forte des schémas I et II 105

2. Vérification des hypothèses d'approximation des convexes et des fonctionnelles 110

2.1. Hypothèses sur K 110
2.2. Hypothèses sur $a(t;u,v)$ 112
2.3. Hypothèses sur $j(t,v)$ 112
2.4. Hypothèses sur $f(t)$ 112

VII. RESOLUTION NUMERIQUE DE QUELQUES PROBLEMES 114

1. Méthodes théoriques 114
1.1. Méthode de minimisation sans contrainte 115
1.2. Méthode de minimisation avec contraintes de bornes 118
1.3. Méthodes de minimisation sous contraintes convexes 121
1.3.1. Méthode duale semi-implicite 121
1.3.2. Méthode duale de double gradient 125
1.4. Méthode de pénalisation par troncature 128

2. Exemples numériques 131
2.1. Problème de Dirichlet 131
2.2. Problème de Neumann 145
2.3. Un problème de climatisation stationnaire 156
2.4. Un problème non différentiable stationnaire 167
.2.5. Le problème de la torsion élasto-plastique 173
2.6. Le problème de la torsion élasto-plastique en évolution 181
2.7. Un problème de frottement évolutif 194
2.8. Un problème parabolique à contraintes réparties 205
2.9. Un problème hyperbolique à contraintes réparties d'ordre 0 226
2.10. Un problème hyperbolique à contraintes réparties d'ordre 1 234
2.11. Conclusion générale 240

BIBLIOGRAPHIE

NOTATIONS

$\mathbb{R}$ corps des réels.

Ω ouvert borné de $\mathbb{R}^N$.

$L^2(\Omega)$ espace des fonctions de carré sommable sur Ω .

$H^1(\Omega)$ espace des fonctions de $L^2(\Omega)$ dont les dérivées à l'ordre 1 appartiennent à $L^2(\Omega)$.

$\mathcal{D}(\Omega)$ espace des fonctions indéfiniment dérivables à support compact dans Ω

$H_o^1(\Omega)$ fermeture de $\mathcal{D}(\Omega)$ dans $H^1(\Omega)$.

$H^{-1}(\Omega)$ dual de $H_o^1(\Omega)$.

$\mathcal{C}^i(\Omega)$ espace des fonctions continues sur Ω dont toutes les dérivées jusqu'à l'ordre i sont continues.

$L^\infty(\Omega)$ espace des fonctions essentiellement bornées sur Ω .

$L^2(0,T ; V)$ espace des fonctions de $[0,T] \subset \mathbb{R}$ à valeurs dans un espace normé V, dont la norme dans V est de carré sommable sur $[0,T]$.

$L^\infty(0,T ; V)$ espace des fonctions de $[0,T] \subset \mathbb{R}$ à valeurs dans un espace normé V dont la norme est essentiellement bornée sur $[0,T]$.

$\mathcal{C}^i(0,T ; V)$ espace des fonctions continues et dont les dérivées en $t \in [0,T]$ sont continues de $[0,T]$ dans un espace normé V .

$\mathcal{D}(0,T ; V)$ espace des fonctions indéfiniment dérivables à support compact dans $[0,T]$, qui appliquent $[0,T]$ dans un espace normé V.

§

C H A P I T R E I

EXISTENCE ET UNICITE POUR LES PROBLEMES ELLIPTIQUES

1. FORMULATION ABSTRAITE

On se donne :

(1.1) V un espace de Banach réflexif sur $\mathbb{R}$, muni d'une norme $\| \ \|$
V' le dual topologique de V , le produit scalaire de la dualité est noté $(\ , \ )$

(1.2) K un sous-ensemble convexe, fermé, non vide de V

(1.3) $A : V \longrightarrow V'$ un opérateur linéaire, continu et monotone, i.e.
$(Au - Av, u-v) \geqslant 0 \quad , \quad \forall\, u \in V \quad , \quad \forall\, v \in V$

(1.4) $j(v)$ une fonctionnelle, semi-continue inférieurement (i.e., s.c.i.) sur V et $\not\equiv \pm \infty$ sur K

(1.5) f un élément de V'.

On pose

$$a(u, v) = (Au, v) .$$

Dans ce chapitre, nous allons étudier les deux problèmes :

$$\begin{cases} \text{Trouver } u \in K \text{ tel que} \\ a(u, v-u) + j(v) - j(u) \geqslant (f, v-u) \quad , \quad \forall\ v \in K. \end{cases}$$

$$\begin{cases} \text{Trouver } u \in K \text{ tel que} \\ a(v, v-u) + j(v) - j(u) \geqslant (f, v-u) \quad , \quad \forall\ v \in K\ . \end{cases}$$

Pour simplifier l'écriture, nous notons ces problèmes respectivement par

(1.6) $\quad a(u, v-u) + j(v) - j(u) \geqslant (f, v-u) \quad , \quad \forall\ v \in K \quad , \quad u \in K \quad ,$

(1.7) $\quad a(v, v-u) + j(v) - j(u) \geqslant (f, v-u) \quad , \quad \forall\ v \in K \quad , \quad u \in K \quad .$ ∎

--- § ---

2. RESULTATS PRELIMINAIRES

On démontre sans difficulté les résultats suivants :

Lemme 2.1. Les problèmes (1.6) et (1.7) sont équivalents : toute solution (s'il en existe) de l'un est solution de l'autre et réciproquement ∎

Remarque 2.1.
Le Lemme 2.1 est en défaut si $j(v)$ est seulement quasi-convexe. Le contre-exemple suivant le prouve. Avec $V = \mathbb{R}$, $\Omega = [0,1]$, $a(v,w) = v\,w$, $j(v) = -v^2$, $f = 0$, $K = [0,1]$, l'inéquation (1.6) a pour solution $u = 1$ et l'inéquation (1.7) a pour solutions $u = 0$ et $u = -1$.
Dans ce cas, les inéquations (1.6) et (1.7) ne sont pas équivalentes. ∎

Remarque 2.2.
Si la fonctionnelle $j(v)$ est dérivable, l'inéquation (1.6) est équivalente à :

$$a(u, v-u) + (j'(u), v-u) \geqslant (f, v-u), \quad \forall\, v \in K, \quad u \in K. \quad \blacksquare$$

Lemme 2.2. Si $a(u,v)$ est forme bilinéaire monotone alors $a(v,v)$ est une fonctionnelle convexe sur V. ∎

Lemme 2.3. Sous les hypothèses (1.1)(1.2)(1.4), avec v fixé $\in K$ et deux constantes $\alpha > 0$, $C \geqslant 0$, l'ensemble des $u \in K$ tels que

$$\alpha \| u - v \|^2 \leqslant j(v) - j(u) + C \| v - u \|$$

est un sous-ensemble borné de K. ∎

Remarque 2.3.
Le lemme 2.3 est en défaut si $j(v)$ est seulement quasi-convexe. Par exemple, avec $V = K = \Omega = \mathbb{R}$, $j(v) = v^3$, on a $-\frac{v^3}{|v|} \to +\infty$ si $v \to -\infty$. ∎

Le lemme 2.3 permet de démontrer le lemme suivant :

Lemme 2.4. Sous les hypothèses (1.1) ... (1.5), si K n'est pas borné, et si l'on a l'une des deux hypothèses supplémentaires :
$a(u,v)$ est coercif, i.e.

$$a(v,v) \geqslant \alpha \| v \|^2, \quad \forall\, v \in V, \quad \alpha > 0,$$

ou bien

$j(v)$ est fortement convexe, i.e. $j(v)$ est convexe et

$$j(u + \tfrac{1}{2}(v-u)) \leq j(u) + \tfrac{1}{2}[j(v) - j(u)] - \beta \|v-u\|^2, \ \beta > 0, \forall u, \forall v,$$

il est possible de définir un convexe fermé borné K_R tel que les deux problèmes

$$a(u, v-u) + j(v) - j(u) \geq (f, v-u), \ \forall v \in K, \ u \in K,$$

$$a(u_R, v_R-u_R) + j(v_R) - j(u_R) \geq (f, v_R-u_R), \ \forall v_R \in K_R, \ u_R \in K_R.$$

soient équivalents (toute solution de l'un, s'il en existe, est solution de l'autre et réciproquement). ∎

Lemme 2.5 (unicité). Si $a(u,v)$ est strictement monotone, i.e.

$$a(u-v, u-v) > 0, \ \forall u \in K, \ \forall v \in K, \ u \neq v$$

ou bien si $j(v)$ est strictement convexe, i.e $\forall \theta \in]0,1[$

$$j(u + \theta(v-u)) < j(u) + \theta[j(v)-j(u)], \ \forall u \in K, \forall v \in K, u \neq v$$

et si (1.6) a une solution, alors cette solution est unique. ∎

Avant d'aborder le problème de l'existence de solutions pour les problèmes (1.6)(1.7) nous rappelons le Théorème de Ky-Fan et Sion:

Théorème 2.1 (Ky-Fan et Sion)(1). Soient E et F deux espaces vectoriels topologiques et C, D, deux sous-ensembles convexes et compacts de E et F respectivement.

Soit $L(u,v)$ une fonction numérique définie sur $C \times D$ à valeurs dans $\bar{\mathbb{R}}$,

- quasi-concave et s.c.s.(2) sur C en u, pour v fixé quelconque
- quasi convexe et s.c.i.(3) sur D en v, pour u fixé quelconque,

alors il existe un couple $(u,v) \in C \times D$ tel que

$$(2.1) \qquad L(\tilde{u},v) \leq L(u,v) \leq L(u,\tilde{v}), \quad \forall (\tilde{u},\tilde{v}) \in C \times D. \ ∎$$

---

(1) Cf. BERGE [1], BROWDER [9].

(2) Semi-continue-supérieurement.

(3) Semi-continue-inférieurement.

L'ensemble des couples $(u,v) \in C \times D$ tels que (2.1) est l'ensemble S des points de selle de $L(u,v)$ sur $C \times D$. ■

On démontre sans difficulté (cf. MANGASARIAN et PONSTEIN [1], TREMOLIERES [1]) le théorème suivant :

Théorème 2.2. Si C et D sont deux sous-ensembles convexes et compacts de E et F respectivement et si $L(u,v)$ est une fonction numérique sur $C \times D$,

- s.c.s. en $u \in C$, pour v fixé quelconque $\in D$
- s.c.i. en $v \in D$, pour u fixé quelconque $\in C$

on a :

$$(2.2) \quad \min_{v \in D} \max_{u \in C} L(u,v) = \max_{u \in C} \min_{v \in D} L(u,v) \quad , \quad (= r)$$

si et seulement si :

(2.3) S est non vide. ■

On note R l'ensemble des couples $(u,v) \in C \times D$ tels que

$$(2.4) \quad L(u, v) = r$$

L'ensemble R est appelé l'ensemble des points minimax - maximin. ■

On a le :

Corollaire 2.1 Sous les hypothèses du Théorème 2.1, on a

$$(2.5) \quad S \equiv R \neq \emptyset \, . \quad ■$$

3. EXISTENCE

Avec :

$$L(u,v) = a(u, v-u) + j(v) - j(u) - (f, v-u) \quad , \tag{3.1}$$

$$E = F = V \quad , \tag{3.2}$$

$$C = D = K \quad , \tag{3.3}$$

les hypothèses (1.1)...(1.5), et l'une ou l'autre des hypothèses du Lemme 2.4, il est immédiat de vérifier que le Théorème 2.1 est vrai.
D'où l'existence de $(u_1, u_2) \in K \times K$ tel que :

$$L(v,u_2) \leqslant L(u_1, u_2) \leqslant L(u_1, w), \qquad \forall\ (v, w) \in K \times K \quad .$$

En faisant $v = u_2$, $w = u_1$, on obtient :

$$\begin{aligned} &L(u_1, u_2) = 0 \quad , \\ &L(u_1, w) \geqslant 0 \quad , \quad \forall\ w \in K \ , \\ &L(v, u_2) \leqslant 0 \quad , \quad \forall\ v \in K \ . \end{aligned}$$

ce qui montre que les problèmes (1.6) et (1.7) ont chacun au moins une solution.
Avec les Lemmes 2.1, 2.4, 2.5, on peut alors énoncer :

Théorème 3.1 (existence et unicité). Sous les hypothèses (1.1) ... (1.5), à condition que K soit borné, ou que a(u,v) soit coercif, ou encore que j(v) soit fortement convexe, les inéquations (1.6) et (1.7) ont chacune au moins une solution. De plus, si a(u,v) est strictement monotone (ce qui est le cas si a(v,v) est coercif), ou si j(v) est strictement convexe (ce qui est le cas si j(v) est fortement convexe), les inéquations (1.6) et (1.7) ont une seule et même solution. ∎

Remarque 3.1.
Dans tout ce qui précède, on peut remplacer l'hypothèse de coercivité par la condition : $\exists\ v_0 \in K$ tel que

$$\frac{a(v,\ v-v_0)}{\|v\|} \longrightarrow +\infty \quad \text{si} \| v \| \longrightarrow +\infty \quad . \qquad \blacksquare$$

Remarque 3.2 .

Lorsque A est hermitien (symétrique dans le cas réel), on montre aisément que le problème (1.6) est équivalent à :

$$\begin{cases} \underset{v \in K}{\text{minimiser}} \; J(v) \\ \text{avec} \qquad J(v) = \frac{1}{2}\, a(v,v) + j(v) - (f,v) \; . \end{cases}$$

L'existence résulte alors directement des théorèmes classiques d'existence en optimisation. L'inéquation (1.6) exprime alors une caractérisation particulière du minimum (conditions d'Euler). ∎

Remarque 3.3 .

On établirait aisément le résultat suivant : sous les hypothèses (1.1)(1.2)(1.3)(1.5), K borné et j(v) quasi-convexe et faiblement s.c.i. (qui affaiblit (1.4)), les problèmes (1.6)(1.7) ont chacun au moins une solution. Mais ici, l'équivalence des solutions n'est plus vraie (cf. Remarque 2.1) et la coercivité n'assure pas l'unicité. ∎

Remarque 3.4 .

La méthode que nous venons de donner pour démontrer l'existence repose sur le Théorème de **Ky-Fan et Sion** ; les méthodes classiques font appel au Théorème de point-fixe de **Brouwer et Kakutani** et de leurs généralisations (cf. BERGE [1] p. 118, BROWDER [9] , BREZIS [2]).

Dans ces méthodes, tout revient à montrer que l'application multivoque

$$u \longrightarrow \Gamma u = \{\, v \mid v \in K \;\; , \;\; (Au - f,\, v-u) + j(v) - j(u) \leqslant 0 \,\}$$

possède un point fixe dans K.

Quant à la formulation "Lagrangienne" de l'inéquation variationnelle, elle permet d'établir l'existence pour deux inéquations en même temps. Cette formulation permet aussi (cf. Ch. III) de démontrer l'existence de solutions faibles dans les problèmes d'évolution. Par ailleurs, il est possible d'en déduire des algorithmes concernant la résolution des problèmes non symétriques (cf. TREMOLIERES [4]). ∎

---- § ----

C H A P I T R E II

APPROXIMATION DES PROBLEMES ELLIPTIQUES

On envisage dans ce Chapitre, l'approximation externe des inéquations variationnelles elliptiques.

1. THEORIE

1.1 Cadre fonctionnel

Nous prenons pour cadre théorique de l'approximation les données suivantes :

V, espace de Hilbert , produit scalaire $((\ ,\ ))$, norme $\|\ \|$

V', dual de V , norme $\|\ \|^*$

H , espace de Hilbert , produit scalaire $(\ ,\ )$, norme $|\ |$

H', dual de H , H' identifié à H, ce qui permet de noter encore par $(\ ,\ )$ le produit scalaire de la dualité entre V et V'

$V \subset H$ avec injection continue et densité

Φ , espace de Hilbert , produit scalaire $[\ ,\ ]_\Phi$, norme $[\ ]_\Phi$

$F = H \times \Phi$ muni du produit scalaire et de la norme produit (1)

$\pi : F \longrightarrow H$, première projection canonique.

$\varpi : F \longrightarrow \Phi$, seconde projection canonique.

$\sigma : V \longrightarrow F$, isomorphisme d'extension de V sur $\sigma V \subset F$.

---

(1) Pour simplifier, nous n'introduisons pas d'espace $\overline{H}$, "extension" de H.

$\pi\ \sigma$ = identité

h, un paramètre destiné à tendre vers 0

V_h, espace de Hilbert , produit scalaire $((\ , \ ))$; norme $\| \ \|_h$, (en général, cet espace sera de dimension finie N(h) avec $N(h) \longrightarrow +\infty$ lorsque $h \longrightarrow 0$)

p_h , opérateur linéaire de $V_h \longrightarrow$ F appelé prolongement de V_h à F et nous supposons que p_h vérifie une condition dite de stabilité :
$\| p_h v_h \|_F \leqslant C \| v_h \|_h$, C constante indépendante de h. ∎

1.2. Approximation de K (ou de V lorsque K = V)

On se donne :

| K convexe fermé de V ,
| K_h convexe fermé de V_h .

Nous supposons que le triplet $\{ p_h, K_h, K \}$ est

- consistant, i.e.

(1.1) $$v_h \in K_h \ , \ p_h v_h \longrightarrow v \text{ dans F faible } \Longrightarrow v \in \sigma K \ ,$$

- convergent, i.e.

(1.2) $$\forall\ w \in K, \ \exists\ w_h \in K_h \text{ tel que } \|w_h\|_h \leqslant C \text{ et } p_h w_h \longrightarrow \sigma w \text{ dans F fort.}$$

Avec (1.1)(1.2) nous dirons que K_h approche K. ∎

Remarque 1.1.
Si $\mathcal{K}$ est un ensemble convexe dense dans K, on peut remplacer (1.2) par

(1.2bis) $$\forall \varphi \in \mathcal{K} \ , \exists \varphi_h \in K_h, \ p_h \varphi_h \longrightarrow \sigma \varphi \text{ dans F fort.}$$ ∎

1.3. Approximation de la forme a (u,v)

Etant donnée $a(u,v)$ une forme bilinéaire, continue, coercive sur V, on se donne $a_h(u_h, v_h)$ forme bilinéaire sur V_h telle que

(1.3) $$|a_h(u_h, v_h)| \leqslant C \ . \ \| u_h \|_h \ . \| v_h \|_h \ , \ \forall\ u_h, v_h \in V_h$$

avec C constante $\geqslant 0$ indépendante de h .

(1.4) $$a_h(v_h, v_h) \geqslant \alpha \| v_h \|_h^2 \ , \ \forall\ v_h \in V_h$$

avec α constante > 0 indépendante de h.

Nous supposons que le triplet $\{p_h,\ a_h(\ ,\ ),\ a(\ ,\ )\}$ est

- consistant, i.e.

(1.5) $$v_h \in K_h,\quad p_h v_h \longrightarrow \sigma v \text{ dans F faible} \Rightarrow \liminf a_h(v_h, v_h) \geq a(v,v)$$

- convergent, i.e.

(1.6) $$\left|\begin{array}{l} \forall v \in K, \forall w \in K, \text{ si } v_h \in K_h \text{ et } w_h \in K_h \text{ sont tels que} \\ p_h v_h \longrightarrow \sigma v \text{ dans F faible}, \quad p_h w_h \longrightarrow \sigma w \text{ dans F fort}, \end{array}\right.$$

alors

(1.6.a) $$a_h(v_h, w_h) \longrightarrow a(v, w),$$

(1.6.b) $$a_h(w_h, v_h) \longrightarrow a(w, v).$$ (1)

Avec les hypothèses (1.3)(1.4), la consistance et la convergence, nous dirons que $a_h(\ ,\ )$ approche $a(\ ,\ )$. ■

1.4. Approximation de $j(v)$

On se donne

$$\left|\begin{array}{l} j(v) : V \longrightarrow \mathbb{R}, \text{ convexe, s.c.i., } \not\equiv \pm\infty \text{ sur } K, \\ j_h(v_h) : V_h \longrightarrow \mathbb{R}, \text{ convexe, s.c.i., } \not\equiv \pm\infty \text{ sur } K_h. \end{array}\right.$$

Nous supposons que le triplet $\{p_h,\ j_h, j\}$ est

- consistant, i.e.

(1.7) $$v_h \in K_h,\ p_h v_h \longrightarrow \sigma v \text{ dans F faible} \Rightarrow \liminf j_h(v_h) \geq j(v),$$

- convergent, i.e.

(1.8) $$w \in K,\ w_h \in K_h,\ p_h w_h \longrightarrow \sigma w \text{ dans F fort} \Rightarrow j_h(w_h) \longrightarrow j(w).$$ (2)

Avec la consistance et la convergence, nous dirons que j_h approche j. ■

Proposition 1.1. Sous les hypothèses introduites, avec $v^o \in K$, $v_h^o \in K_h$, $p_h v_h^o \longrightarrow \sigma v^o$ dans F fort, $u_h \in K_h$ et

(1.9) $$\alpha \| u_h - v_h^o \|_h^2 \leq j_h(v_h^o) - j_h(u_h) + M \| u_h - v_h^o \|_h, \quad \forall\, h \text{ avec } \alpha > 0,\ M \geq 0,$$

on a

---

(1) Cette hypothèse est sans objet si l'on a (1.6.a) et si $a_h(\ ,\ )$ et $a(\ ,\ )$ sont symétriques.

(2) Avec (1.7) il suffit d'avoir $\limsup j_h(w_h) \leq j(w)$ pour assurer (1.8).

(1.10) $\| u_h - v_h^o \|_h \leq C$, constante indépendante de h. ∎

Démonstration

Tout d'abord, si $\| u_h - v_h^o \|_h \leq 1, \forall h$, le résultat est trivial. Supposons alors qu'il existe une suite de valeurs de h , $h \to 0$, telle que

(1.10') $\| u_h - v_h^o \|_h \geq 1$ et $\| u_h - v_h^o \|_h \to + \infty$.

Nous allons montrer que

$$\frac{j_h(v_h^o) - j_h(u_h)}{\| u_h - v_h^o \|_h}$$

reste borné. Les conditions (1.9) et(1.10') étant alors incompatibles, nous aurons donc prouvé (1.10). Avec la convexité, nous avons

$$\theta \, [j_h(v_h^o) - j_h(u_h)] \leq j_h(v_h^o) - j_n((1-\theta)v_h^o + \theta u_h)$$

Prenons $\theta = 1 / \| u_h - v_h^o \|_h$ et $\tilde{u}_h = (1-\theta)v_h^o + \theta u_n$.

Alors $$\frac{j_h(v_h^o) - j_h(u_h)}{\| u_h - v_h^o \|_h} \leq j_h(v_h^o) - j_h(\tilde{u}_h) .$$

Fixons h. Soit $C(v_h^o, 1)$ le cercle de centre v_h^o et de rayon 1. Soit u_h^* tel que

$$u_h^* \in C(v_h^o , 1) \cap K \text{ et}$$
$$j_h(u_h^*) \leq j_h(v_h) , \qquad \forall v_h \in C(v_h^o, 1) \cap K .$$

Alors $$\frac{j_h(v_h^o) - j_h(u_h)}{\| u_h - v_h^o \|_h} \leq j_h(v_h^o) - j_h(u_h^*) = B_h .$$

Montrons maintenant que les B_h sont uniformément bornés. Supposons $h \to 0$ et $B_h \to +\infty$. Comme $\| p_h v_h \|_F \leq \| v_h \|_h \leq 1$, $\forall v_h \in C(v_h^o, 1) \cap K$, alors

$p_h \, C(v_h^o, 1) \cap p_h K_h$ est faiblement compact et il existe une sous-suite u_h^* telle que

$$\left| \begin{array}{l} \limsup \, [j_h(v_h^o) - j_h(u^*_h)] \to + \infty , \\ p_h u_h^* \to \sigma u \text{ dans } F \text{ faible} , \text{ avec } u \in K \text{ (hyp. (1.1)).} \end{array} \right.$$

Donc, d'après(1.8) on a $j(v^o) - \liminf j_h(u^*_h) \to +\infty$ et d'après la consistance (1.7) on a

$$j(v^o) - \liminf j_h(u_h^*) \leqslant j(v^o) - j(u) < + \infty .$$

Il y a donc contradiction. En conséquence $\| u_h - v_h^o \|_h \leqslant C$, constante indépendante de h. ∎

Remarque 1.2.

Avec les hypothèses suivantes : j est continu, j_h est uniformément continu sur V_h [1], $\| p_h v_h \|_F \geqslant \| v_h \|_h$, $\forall \varphi \in \mathcal{K}$ sous-ensemble dense dans K, $\exists \varphi_h \in V_h$ tel que $p_h \varphi_h \longrightarrow \sigma\varphi$ dans F fort et $j_h(\varphi_h) \longrightarrow j(\varphi)$, alors la condition de convergence (1.2) est vérifiée. ∎

1.5. Approximation de f

On se donne :

$f \in V'$,

$f_h \in V_h'$ (V_h' étant le dual de V_h ; la dualité V_h, V'_h étant définie par $(,)_h$; on note $\| \quad \|_h^*$ la norme duale canonique).

On suppose que

(1.11) $\|f_h\|_h^* \leqslant C$, constante indépendante de h

et que l'on a consistance et convergence, c'est-à-dire :

(1.12) $v_h \in K_h$, $p_h v_h \longrightarrow \sigma v$ dans F faible $\Longrightarrow (f_h, v_h)_h \longrightarrow (f,v)$. ∎

1.6. Problèmes approchés

Avec les données et hypothèses précédentes, on considère le problème approché de (1.6) :

(1.13) $a_h(u_h, v_h - u_h) + j_h(v_h) - j_h(u_h) \geqslant (f_h, v_h - u_h)_h , \forall v_h \in K_h, u_h \in K.$

En vertu des hypothèses faites, ce problème a une solution unique. Nous allons voir que si h est assez petit, la solution u_h approche convenablement (dans un sens que nous préciserons) la solution u. ∎

---

(1) ou éventuellement sur K_h, ce qui impose de prendre $\varphi_h \in K_h$.

1.7. Convergence faible

Théorème 1.1. Avec les hypothèses faites précedemment[1] on a

$$p_h u_h \longrightarrow \sigma u \text{ dans } F \text{ faible.}$$

Démonstration

La démonstration se fait en trois temps :

a) Montrer que $\|u_h\|_h$ est borné indépendamment de h

Soient $v^0 \in K$ et $v_h^0 \in K_h$ tel que $\|v_h^0\| \leq C$ et $p_h v_h^0 \longrightarrow \sigma v^0$ dans F fort (hypothèse (1.2)). L'inéquation (1.13) donne :

$$a_h(u_h - v_h^0 , u_h - v_h^0) \leq j_h(v_h^0) - j_h(u_h) + (f_h , u_h - v_h^0)_h + a_h (v_h^0 , v_h^0 - u_h) .$$

D'après la continuité uniforme (1.3), la coercivité uniforme (1.4), et (1.11) il vient

$$\alpha \| u_h - v_h^0\|_h^2 \leq j_h(v_h^0) - j_h(u_h) + [\|f_h\|_h^* + C\|v_h^0\|_h] - \|u_h - v_h^0\|_h .$$

Ce qui entraîne (d'après la Proposition 1.1)

$$\|u_h - v_h^0\|_h \leq C \quad , \text{ constante indépendante de } h .^{(2)}$$

D'où l'on obtient aisément, puisque $\|v_h^0\| \leq C$,

$$\|u_h\| \leq C \quad , \text{ constante indépendante de } h .$$

b) Montrer que $p_h u_h$ est borné dans F indépendamment de h

Cela résulte directement du point a) et de la stabilité de p_h (cf. § 1.1).

c) Montrer qu'il existe une sous-suite $(u_{h'})$ telle que $p_{h'} u_{h'} \longrightarrow \sigma u$ dans F faible

Puisque $(p_h u_h)$ est borné dans F, il existe $v \in F$ et une sous-suite $(u_{h'})$ de (u_h) telle que $p_{h'} u_{h'} \longrightarrow v$ dans F faible. D'après la consistance de $\{p_h , K_h, K\}$ (1.1) nous avons $v \in \sigma K$ et nous pouvons écrire $v = \sigma u$ avec $u \in K$. Nous allons montrer que u est solution de (1.6). Soit w quelconque $\in K$. D'après la convergence de $\{p_h, K_h, K\}$ nous pouvons considérer une suite

---

(1) Toutes les hypothèses faites depuis le début du Ch. II sont utilisées, à l'exception toutefois de (1.6b).

(2) D'une ligne sur l'autre, la lettre C peut désigner des constantes différentes.

$(w_{h'})$ telle que $w_{h'} \in K_{h'}$, $p_{h'}\, w_{h'} \longrightarrow \sigma w$ dans F fort.

On a alors

$$a_{h'}(u_{h'}, u_{h'}) + j_{h'}(u_{h'}) \leqslant a_{h'}(u_{h'}, w_{h'}) + j_{h'}(w_{h'}) + (f_{h'}, u_{h'})_{h'} - (f_{h'}, w_{h'})_{h'} .$$

En utilisant les hypothèses (1.5) (1.7) (1.6a) (1.8) (1.12) on peut passer à la limite et l'on obtient

$$a(u,\ w - u) + j(w) - j(u) \geqslant (f,\ w - u) .$$

Ceci étant vrai $\forall\, w \in K$, c'est que u est bien solution de (1.6). Comme cette solution est unique, u est donc point limite de toute sous-suite extraite de (u_h). En conséquence, c'est la suite $(p_h u_h)$ qui tend toute entière faiblement vers u . ∎

1.8. Convergence forte

Théorème 1.2.
Avec les hypothèses faites précédemment (1), on a $p_h\, u_h \longrightarrow \sigma u$ dans F fort. ∎

Démonstration

D'après la convergence de $\{p_h, K_h, K\}$ (1.2) nous pouvons considérer une suite (v_h) , $v_h \in K_h$, telle que $p_h\, v_h \longrightarrow \sigma u$ dans F fort. Posons

$$X_h = a_h\,(u_h - v_h\ ,\ u_h - v_h) .$$

En développant, nous avons

$$X_h = a_h(u_h\ ,\ u_h) - a_h(u_h\ ,\ v_h) - a_h(v_h\ ,\ u_h) + a_h(v_h\ ,\ v_h) .$$

Mais u_h est solution de (1.13) donc

$$a_h(u_h\ ,\ u_h) \leqslant a_h(u_h\ ,\ v_h) + j_h(v_h) - j_h(u_h) + (f_h\ ,\ u_h)_h - (f_h\ ,\ v_h)_h \quad ,$$

et en portant dans X_h il vient

$$X_h \leqslant j_h(v_h) - j_h(u_h) + (f_h\ ,\ u_h)_h - (f_h\ ,\ v_h)_h - a_h(v_h\ ,\ u_h) + a_h(v_h\ ,\ v_h) .$$

En utilisant la coercivité uniforme et les hypothèses (1.8) (1.7) (1.13) (1.6b)

---

(1) Les hypothèses sont les mêmes que pour le Théorème 1.1 avec en plus l'hypothèse (1.6b).

(1.6a) il vient

$$\| u_h - v_h \|_h \longrightarrow 0 .$$

Puis, d'après la stabilité de p_h ,

$$\| p_h u_h - p_h v_h \|_F \longrightarrow 0 .$$

Mais

$$\| p_h u_h - \sigma u \|_F \leq \| p_h u_h - p_h v_h \|_F + \| p_h v_h - \sigma u \|_F$$

et comme par hypothèse $p_h v_h \longrightarrow \sigma u$ dans F fort, on déduit

$$\| p_h u_h - \sigma u \|_F \longrightarrow 0$$

ce qui établit le Théorème 1.2. ∎

1.9. Définition fréquente du problème discrétisé : extension à F

On suppose que l'on peut "étendre" $a(u,v)$, $j(v)$, (f,v) à F (ce qui est possible si σV (1) est fermé dans F) en définissant $\tilde{a}(u,v)$, $\tilde{j}(v)$, $(\tilde{f},v)$ (2) sur F tels que

$$\left| \begin{array}{lll} \tilde{a}(\sigma u, \sigma v) = a(u,v) & , & \forall u,v \in V \\ \tilde{j}(\sigma v) = j(v) & , & \forall v \in V \\ (\tilde{f}, \sigma v) = (f,v) & , & \forall v \in V \end{array} \right. .$$

On suppose que $\tilde{a}(,)$ a les mêmes constantes de continuité C et de coercivité α que $a(,)$, et que $\tilde{j}$ reste convexe et s.c.i. sur F .
On définit (p_h étant supposé injectif) :

$$\left| \begin{array}{l} \| v_h \|_h = \| p_h v_h \|_F \quad , \forall v_h \in V_h \\ ((v_h, w_h))_h = ((p_h v_h , p_h w_h))_F \quad , \quad \forall v_h , w_h \in V_h \\ a_h(v_h , w_h) = \tilde{a}(p_h v_h , p_h w_h) \quad , \quad \forall v_h , w_h \in V_h \\ j_h(v_h) = \tilde{j}(p_h v_h) \quad , \quad \forall v_h \in V_h \\ (f_h , v_h)_h = (\tilde{f} , p_h v_h) \quad . \quad \forall v_h \in V_h \end{array} \right.$$

En ce qui concerne l'approximation de $a(u,v)$ les conditions (1.3) (1.4) (1.5) (1.6) sont alors vérifiées trivialement. Ainsi, la consistance (1.5) résulte

(1) ou bien σK .
(2) On désigne encore par $(,)$ la dualité entre F et F'.

directement de la continuité et de la convexité de $\tilde{a}(v,v)$. Pour démontrer par exemple (1.6a) on procède ainsi : on considère $u_\varepsilon \longrightarrow u$ dans F faible, $v_\varepsilon \longrightarrow v$ dans F fort et l'on doit montrer que $\tilde{a}(u_\varepsilon , v_\varepsilon) \longrightarrow a(u,v)$.
Pour cela, on écrit

$$\tilde{a}(u_\varepsilon , v_\varepsilon) = \tilde{a}(u_\varepsilon , v) + \tilde{a}(u_\varepsilon , v_\varepsilon - v) .$$

Du fait de la linéarité et de la continuité de $\tilde{a}(u_\varepsilon , v)$ on a

$$\tilde{a}(u_\varepsilon , v) \longrightarrow a(u,v) .$$

Par ailleurs

$$\left| \tilde{a}(u_\varepsilon , v_\varepsilon - v) \right| \leqslant C \left\| u_\varepsilon \right\|_F \cdot \left\| v_\varepsilon - v \right\|_F$$

et le terme de droite tend vers 0 car $u_\varepsilon \longrightarrow u$ dans F faible $\Longrightarrow \left\| u_\varepsilon \right\|_F$ borné. En ce qui concerne j_h, la convexité, la s.c.i. et la consistance (1.7) sont vérifiées dès que $\tilde{j}$ est convexe et s.c.i. En général, la convergence (1.8) n'est vérifiée que si $\tilde{j}$ est continu pour la topologie forte, c'est donc ici la seule propriété à vérifier.
Pour f_h, les propriétés (1.12) (1.13) sont immédiates et n'ont donc pas besoin d'être vérifiées. ■

2. APPROXIMATION EXTERNE

Nous allons définir des méthodes de discrétisation dans le cas où V est un espace de fonctions définies sur un ouvert borné Ω de $\mathbb{R}^N$ de frontière Γ suffisamment régulière. Nous prenons pour référence le problème :

(2.0)
trouver $u \in K$ convexe fermé d'un espace V, $H_0^1(\Omega) \subset V \subset H^1(\Omega)$,
tel que
$$a(u, v-u) + j(v) - j(u) \geqslant (f, v-u) , \quad \forall\, v \in K .$$

On pourra considérer V comme un sous-espace de fonctions $H^1(\Omega)$, nulles sur une partie Γ_o de la frontière Γ. ■

2.1. Discrétisation de Ω

On pose $h = (h_1, \ldots, h_N)$, $h_i > 0$, $i = 1, \ldots, N$ et l'on introduit

- un réseau (ou maillage) $\mathbb{R}_h$:

$$\mathbb{R}_h = \{M \mid M \in \mathbb{R}^N , M = (m_1 h_1, \ldots, m_N h_N) , m_i \in \mathbb{Z} , i = 1, \ldots, N\} \ (1) .$$

(1) $\mathbb{Z}$ est l'espace des entiers positifs et négatifs.

- des pavés centrés en chaque noeud du maillage :

$$\varpi_h^o(M) = \prod_{i=1}^{N} \left[(m_i - \tfrac{1}{2})h_i \, , \, (m_i + \tfrac{1}{2})h_i \right[\, .$$

- des ensembles en forme de croix :

$$\varpi_h^1(M) = \bigcup_{i=1,\ldots,N} \varpi_h^o(M \pm \frac{h_i}{2}) \, ,$$

où l'on note $M \pm \frac{h_i}{2}$ les points $M \pm \frac{h_i}{2} \vec{e}_i$, $\vec{e}_i$ étant le vecteur unitaire du i-ème axe de coordonnée.

- une discrétisation dé Ω :

$$\Omega_h = \Omega_h^1 = \{ M | M \in \mathbb{R}_h \, , \, \varpi_h^1(M) \cap \Omega \neq \emptyset \, , \, \varpi_h^1(M) \cap \Gamma_o = \emptyset \} \quad (2).$$

- les ensembles :

$$\Omega_h^o = \{ M | M \in \mathbb{R}_h \, , \, \varpi_h^o(M) \cap \Omega \neq \emptyset \, , \, \varpi_h^o(M) \cap \Gamma_o = \emptyset \} \, ,$$

$$\overset{o}{\Omega}{}_h^1 = \{ M | M \in \mathbb{R}_h \, , \, \varpi_h^1(M) \subset \Omega \} \, .$$

Remarque 2.1.

Dans le cas où l'on a $H_o^j(\Omega) \subset V \subset H^j(\Omega)$, $j \geq 1$, on définit

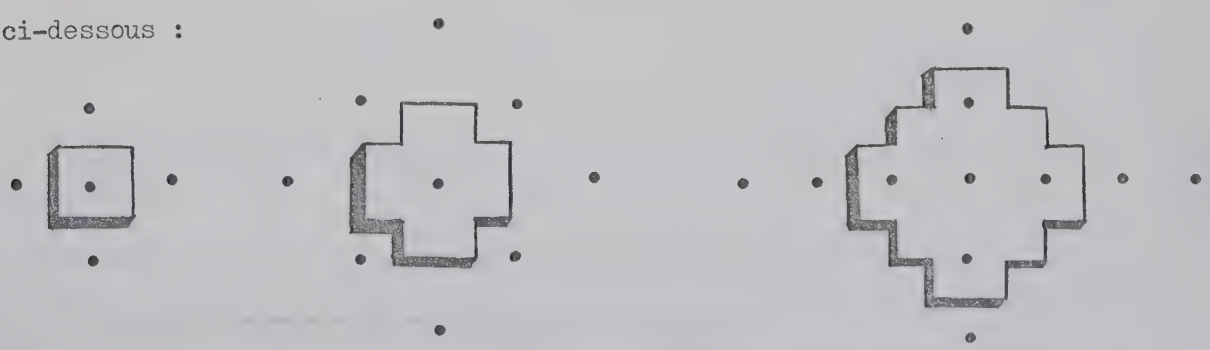
$$\varpi_h^j(M) = \bigcup_{i=1,\ldots,N} \varpi_h^{j-1}(M \pm \frac{h_i}{2})$$

et

$$\Omega_h = \Omega_h^j = \{ M | M \in \mathbb{R}_h, \, \varpi_h^j \cap \Omega \neq \emptyset \, , \, \varpi_h^j(M) \cap \Gamma_o = \emptyset \} \, . \blacksquare$$

Remarque 2.2.

Dans $\mathbb{R}^2$, les ensembles $\varpi_h^o(M)$, $\omega_h^1(M)$, $\omega_h^2(M)$ se représentent comme ci-dessous :



(2) Pour être plus précis, par $\neq \emptyset$ nous entendons "de mesure non nulle" ; de même $= \emptyset$ doit être compris comme "de mesure nulle".

2.2. Discrétisation de V , l'espace V_h et le produit scalaire $((,))_h$

Soit θ_h^M la fonction caractéristique de l'ensemble $\varpi_h^o(M)$. Nous définissons V_h comme l'espace des fonctions en escalier de la forme

$$v_h = \sum_{M \in \Omega_h} v_h^M \cdot \theta_h^M \quad , \quad v_h^M \in \mathbb{R} \, .$$

Nous devons choisir un produit scalaire "proche" de celui de V . Pour cela, nous définissons une approximation de la dérivée d'une fonction v :

$$\nabla_{h_i} v(x) = \frac{1}{h_i} \left[v(x + \frac{h_i}{2} \vec{e}_i) - v(x - \frac{h_i}{2} \vec{e}_i) \right]$$

puis

$$((v_h , w_h))_h = \int_\Omega v_h(x) \, w_h(x) \, dx + \sum_{i=1}^N \int_\Omega \nabla_{h_i} v_h(x) \; \nabla_{h_i} w_h(x) dx$$

$$\|v_h\|_h^2 = \int_\Omega |v_h(x)|^2 \, dx + \sum_{i=1}^N \int_\Omega |\nabla h_i \; v_h(x)|^2 \, dx \, .$$

Il est aisé de voir que l'on définit ainsi un produit scalaire et une norme (cf. CEA [1] , RAVIART [2]).

2.3. L'opérateur de prolongement p_h et la consistance de $\{ p_h , V_h , V \}$

Avec V tel que $H_o^1(\Omega) \subset V \subset H^1(\Omega)$ on pose :

$$F = L^2(\Omega) \times [L^2(\Omega)]^N \, ,$$

$$((\tilde{v},\tilde{w}))_F = \sum_{i=0}^N (v_i , w_i)_{L^2(\Omega)} \, ,$$

avec $\tilde{v} = (v_o , v_1 , \ldots , v_N) \in F$, $\tilde{w} = (w_o , w_1 , \ldots , w_N) \in F$,

$$\|\tilde{v}\|_F = \left(\sum_{i=0}^N |v_i|^2_{L^2(\Omega)} \right)^{\frac{1}{2}} \quad (1),$$

$$p_h : v_h \in V_h \longrightarrow (v_h|_\Omega \; ; \; \nabla_{h_1} v_h|_\Omega \; , \ldots , \; \nabla_{h_N} v_h|_\Omega \;) \in F \, .$$

Nous allons voir que p_h vérifie les propriétés requises :

- p_h est stable. En effet $\|p_h v_h\|_F = \|v_h\|_h$.
- $\{ p_h , V_h , V \}$ est consistant (i.e. $p_h v_h \longrightarrow v$ dans F faible $\Longrightarrow v \in \sigma V$)

---

(1) Lorsqu'il n'y a pas de confusion possible nous omettons d'écrire F en indice inférieur de la norme et du produit scalaire.

Pour la démonstration, on part de $v_h \longrightarrow v_o$ dans $L^2(\Omega)$ faible et $\nabla_{h_i} v_h \longrightarrow v_i$ dans $L^2(\Omega)$ faible, $i = 1,\ldots, N$ et l'on montre que

$v_i = \dfrac{\partial v_o}{\partial x_i}$ (cf. CEA [1] , RAVIART [2]) en sorte que $p_h v_h \longrightarrow v$ dans F faible donne $v \in \sigma V$. ■

2.4. L'opérateur r_h et la convergence de $\{p_h , V_h , V\}$

Il est nécessaire d'introduire un nouvel opérateur afin de vérifier les hypothèses de convergence. Dans le cas où $\Omega_h = \Omega_h^1$, on pose :

$$r_h v = \sum_{M \in \Omega_h} \frac{1}{\text{mes}(\omega_h^o(M))} \left[\int_{\omega_h^o(M)} v(x)\, dx \right] \theta_h^M \quad , \; v \in L^2(\Omega) .$$

On vérifie immédiatement que $r_h v \in V_h$. Par contre si $\Omega_h = \Omega_h^1$ "déborde" sur Ω il faut "prolonger" v, pour autant que cela soit possible, à $\mathbb{R}^N$. C'est la raison d'être de l'Hypothèse de prolongement : on dit qu'un ouvert Ω possède la propriété de m-prolongement si $\forall s$, $0 \leqslant s \leqslant m$, il existe une application linéaire continue

$\pi : u \to \pi u$ de $H^s(\Omega)$ dans $H^s(\mathbb{R}^N)$ telle que $\pi u = u$, p.p. sur Ω . ■

On définit alors :

$$r_h v = \sum_{M \in \Omega_h} \frac{1}{\text{mes}(\omega_h^o(M))} \left[\int_{\omega_h^o(M)} \pi v(x) dx \right] \theta_h^M \quad , \qquad v \in L^2(\Omega)$$

et l'on montre (cf. CEA [1] , RAVIART [2]) que

$$p_h r_h v \longrightarrow \sigma v \text{ dans } F \text{ fort, } \forall\, v \in V \qquad (1)$$

en sorte que, relativement à V , ceci établit l'hypothèse (1.2) avec $w_h = r_h v$.

2.5. Discrétisation de quelques convexes

Nous allons d'abord donner quelques idées générales sur la vérification des hypothèses de consistance et de convergence du triplet $\{p_h , K_h , K\}$.

Vérification de la consistance de $\{p_h , K_h , K\}$:

a) vérifier que $\{ p_h, V_h , V \}$ est consistant,

b) construire une famille $(K\varepsilon)_{\varepsilon>0}$, $\varepsilon \to 0$ de convexes fermés de V telle

(1) Il faut cependant faire une hypothèse particulière sur la régularité de Γ .

que

(2.1) $\bigcap K_\varepsilon = K$, $K_\varepsilon \supset K_{\varepsilon'} \supset K$, $\varepsilon > \varepsilon' > 0$,

c) Montrer que $\forall\ \varepsilon > 0$, $\exists\ h_\varepsilon$ tel que

(2.2) $p_h K_h \subset \sigma K_\varepsilon$, $\forall\ h \leqslant h_\varepsilon$.

Ces trois points assurent la consistance (1.1). En effet, avec

$$p_h v_h \longrightarrow v \quad \text{dans } F \text{ faible},$$

on a $v \in \sigma V$; puis $v_h \in K_h$ et (2.2) entraînent $v \in \sigma K_\varepsilon$. Comme ceci est vrai $\forall \varepsilon > 0$, on a $v \in \sigma K$. ∎

Remarque 2.3.

Les points b) et c) sont inutiles si $p_h K_h \subset \sigma K$. Mais en général ceci n'est vrai que pour les approximations internes. ∎

Remarque 2.4.

Dans certains cas, on a $K = K_o \cap V$ ou K_o est un convexe fermé de H. Alors les points b) et c) sont inutiles si $\pi p_h K_h \subset K_o$. Démontrons le . Soit $v_h \in K_h$ tel que $p_h v_h \longrightarrow \tilde{v}$ dans F faible. Nous allons montrer que $\tilde{v} \in \sigma K$ moyennant l'hypothèse $\pi\ p_h K_h \subset K_o$. D'après la linéarité et la continuité de π , on a :

$$\pi\ p_h v_h \longrightarrow \pi \tilde{v} \quad \text{dans } H \text{ faible.}$$

Comme $\pi p_h K_h \subset K_o$ et que K_o est fermé dans H , on a $\pi\tilde{v} \in K_o$. Or $\tilde{v} \in \sigma V$ d'après a) , d'où $\pi\tilde{v} \in V$. En réunissant les deux résultats trouvés, on obtient $\pi\tilde{v} \in K_o \cap V$ donc $\pi\tilde{v} \in K$, i.e. $\tilde{v} \in \sigma K$. ∎

Vérification de la convergence de $\{ p_h , K_h , K \}$:

d) vérifier que $\{ p_h , V_h , V \}$ est convergeant,

e) construire une famille $(K_\varepsilon)_{\varepsilon>0}$, $\varepsilon \to 0$, ε dépendant éventuellement de h , telle que

(2.3) $\forall\ h$, $\exists \varepsilon$ tel que $\forall\ v_\varepsilon \in K_\varepsilon$ on a $r_h v_\varepsilon \in K_h$,

f) Montrer que $\forall\ \varepsilon>0$, $\forall w\in K$, $\exists\ v_\varepsilon\in K_\varepsilon$ tel que

$$\|w - v_\varepsilon\| \leqslant 0(\varepsilon)\ . \tag{2.4}$$

Les trois points d), e), f) assurent la convergence (1.2). En effet la suite $w_h = r_h\, v_\varepsilon$ répond à la question puisque

$$\|\sigma w - p_h\, r_h\, v_\varepsilon\| \leqslant \|\sigma w - \sigma v_\varepsilon\| + \|\sigma v_\varepsilon - p_h\, r_h\, v_\varepsilon\|\ .$$

On peut remplaçer les conditions e) f) par la suivante :

g) Prendre $w_h = P_{K_h}[\, r_h\, w\,]$ (P_{K_h} étant la projection sur K_h pour la norme $(u_h, v_h)_h = \int_\Omega v_h(x)\, w_h(x)\, dx$) et montrer que

$$\|r_h\, w - w_h\|_h \leqslant \varepsilon_h\ ,\ \varepsilon_h \to 0\ .$$

En effet

$$\|\sigma w - p_h\, w_h\| \leqslant \|\sigma w - p_h\, r_h\, w\| + \|p_h\, r_h\, w - p_h\, w_h\|.\ \blacksquare$$

Remarque 2.5.

Le point g)est immédiat si $r_h\, K \subset K_h$ ■ Dans la suite de ce paragraphe nous étudions la discrétisation et l'approximation de quelques convexes.

Exemple 2.1.

On pose

$$K = \{\, v \mid v\in H_o^1(\Omega)\ ,\ v(x)\geqslant g(x)\ ,\ \text{p.p. sur } \Omega\,\}$$

avec

$$g(x)\in \mathcal{C}\ (\Omega)\ ,\ g(x)\leqslant 0\ ,\ x\in \partial\Omega = \Gamma$$

et l'on définit

$$K_h = \{\, v_h \mid v_h\in V_h\ ,\ v_h(x)\geqslant g_h(x)\ ,\ x\in \omega_h^o(M)\ ,\ M\in \Omega_h\,\}$$

avec

1er cas : $g_h(x) = r_h\, g(x)$, 2ème cas : $g_h(x) = g(M)$, $x\in \omega_h^o(M)$.

Vérifions la consistance (1er et 2ème cas)

L'hypothèse (2.1) est vérifiée avec

$$K_\varepsilon = \{\, v \mid v\in H_o^1(\Omega)\ ,\ v(x)\geqslant g(x) - \varepsilon\ ,\ \text{p.p. sur } \Omega\,\}\ .$$

L'hypothèse (2.2) est immédiate avec h_ε tel que

$$\sup_{\substack{\|x-y\|\leqslant h_\varepsilon \\ x,y\in\Omega}} \{\, |\,|g(x)| - |g_h(y)|\,|\,\} < \varepsilon\ ,$$

et cette condition est licite car $g \in \mathcal{C}(\Omega)$.

Vérifions la convergence.

Dans le 1er cas, avec $g_h(x) = r_h\, g(x)$, on a $r_h(w - g) \geqslant 0$. Autrement dit, $w_h = P_{K_h}\,[r_h\, w] = r_h\, w \geqslant r_h\, g$ et g) est vérifié. Dans le 2ème cas, nous allons vérifier e) et f). Soit :

$$\left|\begin{array}{l} \eta_h = \sum_{M \in \Omega_h} \eta_h^M\ \theta_h^M\ ,\ \eta_h^M \in \mathbb{R}\ ,\quad \theta_h^M \text{ étant la} \\ \\ \text{fonction caractéristique de } \varpi_h^o(M), \text{ avec} \\ \\ r_h(1 + \eta_h^M)\, g(x) = g(M)\ ,\ x \in \varpi_h^o(M)\ ,\ M \in \Omega_h \cdot \end{array}\right.$$

Posons

$$\varepsilon_h = \sup_{M \in \Omega_h} \left|\eta_h^M\right|$$

et

$$K_{\varepsilon_h} = \{\, v \,|\, v \in V\ ,\ v(x) \geqslant g(x)(1 + \eta_h^M)\ ,\ x \in \varpi_h^o(M)\ ,\ M \in \Omega_h\ \}\cdot$$

Alors $\forall\, v \in K_{\varepsilon_h}$, $r_h\, v \in K_h$ et (2.3) est assuré. (Naturellement, comme $g(x) \in \mathcal{C}(\Omega)$, on a $h \longrightarrow 0 \Rightarrow \varepsilon_h \longrightarrow 0$).

Pour le dernier point, étant donnés $v \in K$ et $\varepsilon_h > 0$ quelconque, on pose

$$v_{\varepsilon_h} = v(1 + \varepsilon_h)\cdot$$

Alors

$$\|v - v_{\varepsilon_h}\| = \varepsilon_h \|v\| = 0(\varepsilon_h)$$

et

$$r_h\, v_{\varepsilon_h} \in K_{\varepsilon_h}\ .$$ Le point f) est donc démontré. ∎

Exemple 2.2.

On considère :

$$K = \{\, v \,|\, v \in H_o^1(\Omega)\ ,\ |v(x)| \leqslant d(x,\Gamma)\ ,\ \text{p.p. sur } \Omega\,\}$$

où $d(x,\Gamma)$ est la plus petite distance euclidienne d'un point $x \in \Omega$ à la frontière Γ . On définit

$$K_h = \{\, v_h \,|\, v_h \in V_h\ ,\ |v_h(x)| \leqslant d_h(x)\ ,\ \text{p.p. sur } \Omega\,\}$$

avec

$$\underline{\text{1er cas}} : \quad d_h(x) = r_h \, d(x,\Gamma) \ ,$$

$$\underline{\text{2ème cas}} : \quad d_h(x) = d(M,\Gamma) \ , \quad x \in \omega_h^o(M) \ , \ M \in \Omega_h \ ,$$

$$= 0 \quad \text{ailleurs} \ ,$$

$$\underline{\text{3ème cas}} : \quad d_h(x) = d(M,\Gamma_h) \ ; \quad x \in \omega_h^o(M) \ , \quad M \in \Omega_h \ ,$$

$$= 0 \quad \text{ailleurs}$$

avec $d(\Gamma_h, \Gamma) \longrightarrow 0$ uniformément.

Pour les 1er 2ème cas la démonstration de la convergence et de la consistance est sensiblement équivalente à ce qui a été fait dans l'Exemple 2.1. Dans le 3ème cas il suffit de remarquer que : $K_{(h)} \longrightarrow K$ **i.e.**

$$\forall \varepsilon \ , \ \exists h \text{ tel que } \quad K_{(h)} \subset K \cup \{ v \mid \| v \| \leqslant \varepsilon \}$$

avec

$$K_{(h)} = \{ v \mid v \in H_o^1(\Omega) \ , \ -d(x,\Gamma_h) \leqslant v(x) \leqslant d(x,\Gamma_h) \ , \text{ p.p. sur } \Omega \} \ .$$

Exemple 2.3.

On pose

$$\boxed{K = \{ v \mid v \in H^1(\Omega) \ , \ v(x) \geqslant 0 \ , \text{ p.p. sur } \Gamma' \subset \Gamma \}}$$

et

$$K_h = \{ v_h \mid v_h \in V_h \ , \ v_h(M) \geqslant 0 \ , \ \forall M \in \Gamma'_h \}$$

avec

$$\Gamma'_h = \{ M \in \Omega_h \mid \omega_h^o(M) \cap \Gamma' \neq \emptyset \}$$

et

$$V_h = \{ v_h \mid v_h = \sum_{M \in \Omega_h^1} v_h^M \theta_h^M \ , \ v_h^M \in \mathbb{R} \} \ ,$$

$$\Omega_h^1 = \{ M \mid M \in R_h \ , \ \omega_h^1(M) \cap \Omega \neq \emptyset \} \ .$$

On utilise le résultat suivant (cf. CEA [1] , Théorème 5.3)

Lemme

Si Γ est assez régulière et si $\| v_h \|_h \leqslant C$, constante indépendante de h ; alors il existe une sous-suite de v_h telle que :

$$\left| \begin{array}{l} p_{h'} \, v_{h'} \longrightarrow \sigma v \quad \text{dans } F \text{ faible, avec } v \in H^1(\Omega) \ , \\ \quad v_{h'}|_\Gamma \longrightarrow v|_\Gamma \quad \text{dans } L^2(\Gamma) \text{ faible.} \end{array} \right. \blacksquare$$

Nous allons vérifier la consistance. Soit $v_h \in K_h$, tel que

$$\left| \begin{array}{l} p_h\, v_h \longrightarrow \sigma v \quad \text{dans } F \text{ faible, avec } v \in H^1(\Omega) \ , \\ v_h|_\Gamma \longrightarrow v|_\Gamma \quad \text{dans } L^2(\Gamma) \text{ faible.} \end{array} \right.$$

Nous devons vérifier que $v \in K$. Pour cela on note que

$$\tilde{K}_h = \{ \tilde{v}_h \,|\, \tilde{v}_h \in L^2(\Gamma') \ , \ \tilde{v}_h(x) = v_h(M) \ , \ M \text{ tel que : } x \in \varpi_h^0(M) \ , \ v_h(M) \in K_h \}$$

est un convexe fermé de $L^2(\Gamma')$. De plus

$$\tilde{K}_h \subset \tilde{K} = \{ v \,|\, v \in L^2(\Gamma) \ , \ v \geqslant 0 \ , \text{ p.p. sur } \Gamma \}$$

et $\tilde{K}$ est un convexe fermé de $L^2(\Gamma)$, d'où l'on déduit que $v|_\Gamma \geqslant 0$, p.p. en $x \in \Gamma$. Démontrons la convergence.

Rappelons qu'avec $V = H^1(\Omega)$, V_h est défini sur Ω_h^1 , Ω_h^1 "débordant" partout sur Ω , et que l'opérateur r_h suppose implicitement le prolongement des fonctions $w \in V = H^1(\Omega)$ à $\tilde{V} = H^1(\mathbb{R}^N)$.

Avec $w \in V$, il est possible que $r_h\, w$ soit <0 sur Γ' . En conséquence, $r_h\, w$ peut ne pas appartenir à K_h . Dans ce qui suit, prenant $w \in K$, nous allons construire une suite $w_h \in K_h$ telle que $p_h\, w_h \longrightarrow \sigma w$ dans F fort.

Nous posons

$$w^+ = \sup \{ 0 \ , \ w \} ,$$

$$w^- = \inf \{ 0 \ , \ w \} .$$

Alors $w^+ \in H^1(\Omega)$, $w^- \in H^1(\Omega)$. Nous supposons que ces fonctions sont implicitement prolongées à $H^1(\mathbb{R}^N)$. On a alors

$$p_h\, r_h\, w^+ \longrightarrow \sigma w^+ \quad \text{dans } F \text{ fort} .$$

Par ailleurs

$$w \in K \Longrightarrow w^- \in V^- = \{ v \,|\, v \in H^1(\Omega) \ , \ v = 0 \text{ sur } \Gamma' \}$$

en conséquence, notant V_h^- l'espace des fonctions en escalier définies sur

$$\{ M \,|\, \varpi_h^1(M) \cap \Omega \neq \emptyset \quad , \quad \varpi_h^1(M) \cap \Gamma' = \emptyset \ \} ,$$

r_h^- l'opérateur de restriction de V^- à V_h^- et p_h^- le prolongement de V_h^- à $F = L^2(\Omega) \times L^2(\Omega)^N$, on a

$$\left| \begin{array}{l} p_h^- r_h^- w^- \longrightarrow \sigma w^- \quad \text{dans } F \text{ fort}, \\ r_h^- w^- \in K_h . \end{array} \right.$$

On remarquera que $V_h^- \subset V_h$ et que $p_h v_h^- = p_h^- v_h^-$, $\forall\, v_h^- \in V_h^-$. Posons

$$w_h = r_h w^+ + r_h^- w^- .$$

Alors

$$w_h \in K_h \quad \text{et} \quad p_h w_h \longrightarrow \sigma w^+ + \sigma w^- \quad \text{dans } F \text{ fort.} \quad \blacksquare$$

Exemple 2.4.

On considère

$$K = \{ v \mid v \in H_o^1(\Omega) , \ | \operatorname{grad} v(x)|^2 \leq 1 , \text{ p.p. sur } \Omega \}$$

Nous adopterons ici une discrétisation de la dérivée différente de celle qui a été donnée. La différence apparait au voisinage de la frontière de Ω . Nous notons encore par ∇_{h_i} la différence finie avec la modification frontière qui suit :

Soit $x \in \Omega$. Définissons $M_i^- \in R_h$ tel que $\omega_h^o(M_i^- + \frac{1}{2} h_i) \ni x$, et $M_i^+ = M_i^- + h_i$

Soit V_h l'espace discrétisé de $H_o^1(\Omega)$. Soit $v_h(x) \in V_h$. On pose :

$$\alpha) \quad \nabla_{h_i} v_h(x) = \frac{v_h(M_i^+) - v_h(M_i^-)}{h_i} \quad \text{si } M_i^- \in \Omega_h^1 \text{ et } M_i^+ \in \Omega_h^1 , \tag{1}$$

$$\beta) \quad \nabla_{h_i} v_h(x) = \frac{v_h(M_i^+) - v_h(\widetilde{M}_i^-)}{|M_i^+ - \widetilde{M}_i^-|} \quad \text{si } M_i^- \notin \Omega_h^1 \text{ et } M_i^+ \in \Omega_h^1$$

où $\widetilde{M}_i^-$ est le point de Γ le plus proche de M_i^- dans la direction $-\vec{e}_i$,

$$\gamma) \quad \nabla_{h_i} v_h(x) = \frac{v_h(\widetilde{M}_i^+) - v_h(M_i^-)}{|\widetilde{M}_i^+ - M_i^-|} \quad \text{si } M_i^- \in \Omega_h^1 \text{ et } M_i^+ \notin \Omega_h^1 \tag{2}$$

où $\widetilde{M}_i^+$ est le point deΓ le plus proche de M_i^+ dans la direction $+\vec{e}_i$,

$$\delta) \quad \nabla_{h_i} v_h(x) = 0 \quad \text{si } M_i^- \notin \Omega_h^1 \text{ et } M_i^+ \notin \Omega_h^1 .$$

---

(1) ici, avec $V = H_o^1(\Omega)$, on a $\Omega_h^1 = \overset{o}{\Omega}{}_h^1$.

(2) Avec $v_h \in V_h$ on a $v_h(\widetilde{M}_i^+) = 0$, $v_h(\widetilde{M}_i^-) = 0$.

On définit alors

$$K_h = \{ v_h \mid v_h \in V_h \ , \ \sum_{i=1}^{N} \mid \nabla_{h_i} \ v_h(x) \mid^2 \leq 1 \ , \text{ p.p. sur } \Omega \ \} .$$

On admet le résultat suivant (dont la démonstration est sensiblement équivalente à celle qui est donnée dans CEA [1] ou RAVIART [2] au sujet de l'approximation par les différences finies usuelles) :

Lemme

Avec $\varphi \in \mathcal{C}^2(\bar{\Omega}) \cap H_o^1(\Omega)$ et :

$$\left| \begin{array}{l} \varphi_h(x) = \begin{cases} \varphi(M) & \text{si} \quad x \in \varpi_h^{\circ}(M) \quad \text{et} \quad M \in \Omega_h^1 = \mathring{\Omega}_h^1 \ , \\ 0 & \text{ailleurs} \ , \end{cases} \\ \\ p_h \ \varphi_h = (\ \varphi_h|_{\Omega} \ ; \ \nabla_{h_i} \ \varphi_h|_{\Omega} \ , \ldots , \ \nabla_{h_N} \ \varphi_h|_{\Omega} \ ) \end{array} \right.$$

où ∇_{h_i} est définie comme précédemment ,

$\mathcal{M}_h$ = sous-ensemble des points de $\mathring{\Omega}_h^1$ dont l'un au moins des points voisins $\notin \mathring{\Omega}_h^1$.

alors, avec les hypothèses :

$$\left| \begin{array}{l} \Gamma \text{ suffisamment régulière} \ , \\ \\ t_h = \sup_{x \in \Gamma} d(x, \mathcal{M}_h) \longrightarrow 0 \ ; \end{array} \right.$$

on a le résultat

$$p_h \ \varphi_h \longrightarrow \sigma \varphi \text{ dans } F \text{ fort.} \quad \blacksquare$$

Nous admettrons une dernière hypothèse :

$$\mathcal{C}^2(\bar{\Omega}) \cap K \text{ dense dans } K \ ^{(1)} . \quad \blacksquare$$

Dans le cas $\alpha)$, $\beta)$, $\gamma)$ un développement de Taylor au 1er ordre donne

$$\nabla_{h_i} \ \varphi_h(x) = \frac{\partial \varphi}{\partial x_i} \ (M_i^- + \xi_i)$$

avec, selon les cas,

$$\xi_i \in [M_i^- , M_i^+] , \quad \xi_i \in [\widetilde{M}_i^- , M_i^+] \text{ ou } \xi_i \in [M_i^- , \widetilde{M}_i^+]$$

et dans tous les cas

$$| \xi_i | \leq | t_h + h | .$$

(1) cf. BOURGAT [1] , GOURSAT [1] p.49 , BREZIS-STAMPACCHIA [1] . GLOWINSKI-LIONS-TREMOLIERES [1] .

Avec l'hypothèse faite sur t_h , il vient $h \to 0 \Rightarrow \xi_i \longrightarrow 0$.

On peut écrire

$$\nabla_{h_i} \varphi_h(x) = \frac{\partial \varphi}{\partial x_i}(x) \cdot \left[\frac{\partial \varphi}{\partial x_i} (M_i + \xi_i) - \frac{\partial \varphi}{\partial x_i} (x) \right].$$

Soit

$$\alpha_h = \sup \left\{ \left| \frac{\partial \varphi}{\partial x_i} (x) - \frac{\partial \varphi}{\partial x_i} (y) \right| \ ; \ i=1,\dots, N \ , \| x - y \| \leqslant t_h + h \ , \ x, y \in \Omega \right\}.$$

Comme $\varphi \in \mathcal{C}^1(\bar{\Omega})$ alors $h \to 0 \Rightarrow \alpha_h \longrightarrow 0$. En outre $\varphi \in K$ donc

$$| \text{grad } \varphi(x) |^2 = \sum_{i=1}^{N} \left| \frac{\partial \varphi}{\partial x_i} (x) \right|^2 \leqslant 1 \quad , \forall \ x \in \Omega$$

et

$$\sum_{i=1}^{N} | \nabla_{h_i} \varphi_h(1) |^2 \leqslant (1 + \varepsilon) + n \alpha_h^2 (1 + \frac{1}{\varepsilon}) \ , \ \varepsilon \text{ quelconque} > 0 \quad (1)(2)$$

Posons

$$\varepsilon = \alpha_h ,$$

$$\theta_h^2 = (1 + \alpha_h) (1 + n \alpha_h),$$

alors

$$\sum_{i=1}^{N} | \nabla_{h_i} \varphi_h(x) |^2 \leqslant \theta_h^2 \ , \text{ et } \theta_h \longrightarrow 1 \text{ lorsque } h \longrightarrow 0 .$$

En posant

$$\psi_h = \frac{\varphi_h}{\theta_h}$$

il vient

$$\sum_{i=1}^{N} | \nabla_{h_i} \psi_h(x) |^2 \leqslant 1 .$$

On obtient alors

$$\| \sigma w - p_h \psi_h \| \leqslant \| \sigma w - \sigma \varphi \| + \| \sigma \varphi - p_h \varphi_h \| + \| p_h \varphi_h - p_h \psi_h \| \leqslant$$

$$\leqslant \varepsilon_1 \text{ (densité)} + \varepsilon_2 \text{ (convergence)} + \varepsilon_3$$

avec

$$\varepsilon_3 = \| p_h \varphi_h - p_h \frac{\varphi_h}{\theta_h} \| \leqslant (1 - \frac{1}{\theta_h}) \| p_h \varphi_h \| \leqslant (1 - \frac{1}{\theta_h}) (\varepsilon_2 + \| \varphi \|)$$

---

(1) si le cas δ) apparait la majoration est évidemment valable.

(2) On utilise la majoration $|a+b|^2 \leqslant (1+\varepsilon) a^2 + (1+\frac{1}{\varepsilon}) b^2$, $\varepsilon > 0$.

expression dans laquelle $h \to 0 \Longrightarrow \theta_h \to 1$ donc $\varepsilon_3 \to 0$.

Finallement, $\forall\, w \in K$, nous venons de montrer que l'on peut construire une suite $\psi_h \in K_h$ telle que $p_h \psi_h \to \sigma w$ dans F fort et ceci démontre la propriété de convergence. ∎

2.6. Discrétisation de $a(u,v)$

Le plus souvent $a(u,v)$ a l'une des trois formes :

$$\left|\begin{array}{l} a(u,v) = \int_\Omega \text{grad } u \text{ grad } v \, dx \text{ , avec } u,\, v \in H_o^1(\Omega), \\ a(u,v) = \int_\Omega uv dx + \int_\Omega \text{grad } u \text{ grad } v \, dx \text{ , avec } u,v \in H^1(\Omega), \\ a(u,v) = \int_\Omega uv dx + \int_\Omega \text{grad } u \text{ grad } v \, dx + \int_\Gamma t_o \; u|_\Gamma \cdot v|_\Gamma \, d\sigma \text{ , } t_o \text{ borné} \end{array}\right.$$

A titre d'exemple nous étudions la seconde expression.

On définit

$$\left|\begin{array}{l} \tilde{u} = (u_o, u_1, \ldots, u_N) \in L^2(\Omega)^{N+1}, \\ \tilde{v} = (v_o, v_1, \ldots, v_N) \in L^2(\Omega)^{N+1}, \\ \tilde{a}(\tilde{u},\tilde{v}) = \int_\Omega u_o v_o \, dx + \ldots + \int_\Omega u_N v_N \, dx \, . \end{array}\right.$$

Alors (cf. § 1.9)

$$\tilde{a}(\sigma u, \sigma v) = a(u,v) \, , \quad \forall u \in V, \quad \forall v \in V \, , \quad V = H^1(\Omega) \, .$$

Puis l'on prend

$$a_h(u_h \, , \, v_h) = \tilde{a}(p_h u_h \, , \, p_h v_h) \quad , \quad u_h \in V_h \, , \, v_h \in V_h \, .$$

Toutes les hypothèses du paragraphe 1.3 sont alors vérifiées trivialement et $a_h(u_h \, , \, v_h)$ approche $a(u,v)$. ∎

2.7. Discrétisation de $j(v)$

Nous donnons quelques exemples :

Exemple 2.5.

On considère

$$\boxed{j(v) = \int_\Omega |v(x)| \, dx} \quad , \quad v \in V \quad , \quad H_o^1(\Omega) \subset V \subset H^1(\Omega) \, , \, \Omega \text{ borné.}$$

On définit

$$\tilde{j}(v) = \int_\Omega |v(x)| \, dx \quad , \quad v \in L^2(\Omega),$$

$$j_h(v_h) = \tilde{j}(p_h v_h).$$

La convexité de $\tilde{j}$ est immédiate. De plus

$$\tilde{j}(v) = \int_\Omega |v(x)|\,dx \leqslant \text{mes } \Omega \left(\int |v(x)|^2\,dx\right)^{\frac{1}{2}}$$

d'où l'on déduit que $\tilde{j}(v)$ est fortement continu sur F. Il résulte (cf. paragraphes 1.4. et 1.9.) que $j_h(v_h)$ approche $j(v)$. ∎

Exemple 2.6.

On considère

$$\boxed{j(v) = \int_{\Gamma'} |v(x)|\,d\sigma}, \quad v \in V \quad , \quad H_0^1(\Omega) \subset V \subset H^1(\Omega)$$

où Γ' est une partie quelconque de la frontière de Ω. On suppose que la frontière $\Gamma = \partial\Omega$ est régulière et que la partie Γ' est bornée. On a (cf. CEA [1] p. 43)

$$\|\gamma u\|^2_{L^2(\Gamma)} \leqslant \varepsilon \sum_{i=1}^{N} \left\| \frac{\partial u}{\partial x_i} \right\|_{L^2(\Omega)} + C(\varepsilon) . \|u\|^2_{L^2(\Omega)}$$

où $C(\varepsilon)$ est une constante > 0, indépendante de u et où γu est la trace de u sur Γ.

Cette inégalité entraîne que $j(v)$ est (fortement) continu. La convexité est par ailleurs immédiate. Mais ici, on ne peut pas définir j_h à partir de $\tilde{j}$ car $\tilde{j}(v)$ avec $v \in L^2(\Omega)$ peut ne pas être défini. On ne peut pas non plus poser directement $j_h(v_h) = j(v_h)$ car v_h peut ne pas être défini sur Γ. On prolonge alors v_h de façon que v_h soit défini partout et l'on pose $j_h(v_h) = j(v_h)$

Nous allons démontrer que les hypothèses introduites à la Remarque 1.3 sont vérifiées :

* j_h est uniformément continu. Pour cela on utilise l'inégalité discrète analogue à (2.5) (cf. CEA [1] Théorème 4.2) d'où l'on déduit

$$\|v_h - w_h\|_h < \varepsilon \Longrightarrow \int_{\Gamma'} |v_h - w_h|^2\,d\sigma \leqslant 0(\varepsilon)$$

et

$$\left| \int_{\Gamma'} |v_h|\,d\sigma - \int_{\Gamma'} |w_h|\,d\sigma \right| \leqslant 0(\varepsilon).$$

Ce qui prouve l'uniforme continuité.

* $\|p_h v_h\|_F \geqslant \|v_h\|_h$. En effet $\|p_h v_h\|_F = \|v_h\|_h$ (cf. paragraphe 2.3).

* $\forall \varphi \in V$ dense dans V , $\exists \varphi_h$ tel que $p_h \varphi_h \to \sigma\varphi$ et $j_h(\varphi_h) \to j(\varphi)$.

Cela est immédiat en prenant $V = \mathcal{C}^2(\bar{\Omega})$, $\varphi \in V$ et

$$\varphi_h = \sum_{M \in \Omega_h^o} \varphi(M) \cdot \theta_h^M \ .$$

En effet l'on démontre sans difficulté que

$$p_h \ \varphi_h \longrightarrow \sigma\varphi \quad \text{dans } F \text{ fort (en fait la convergence est uniforme).}$$

La convergence (1.8) est donc assurée. ∎

Exemple 2.7.

On considère

$$\boxed{j(v) = \int_\Omega g \, |\text{grad } v| \, dx} \quad , \quad v \in V \quad , \quad H_o^1(\Omega) \subset V \subset H^1(\Omega)$$

où Ω est borné et de frontière Γ régulière, et $g \in L^\infty(\Omega)$.

On pose

$$\tilde{j}(v) = \int_\Omega g\Big(\sum_{i=1}^N v_i^2\Big)^{\frac{1}{2}} \, dx$$

avec

$$v = (v_o, v_1, \ldots, v_N) \in L^2(\Omega)^{N+1} \ .$$

On définit

$$j_h(v_h) = \tilde{j}(p_h \, v_h) \ .$$

Comme $\tilde{j}(v)$ est convexe et continu sur F, on déduit immédiatement la consistance et la convergence de j_h (cf. paragraphe 1.9). ∎

2.8. Discrétisation de f

Nous allons chercher à définir une approximation f_h de f.

Supposons (ce qui est le cas général) $f \in V'$. Si σV est fermé dans F, σV est alors un espace de Hilbert pour le produit scalaire $((,))_F$. Par ailleurs $L(\sigma v) = (f,v)$ est forme linéaire continue sur σV. Il est alors possible de définir (d'après le Théorème de Riesz) un élément $\tilde{f} \in \sigma V$ tel que

$$(f,v) = ((\tilde{f}, \sigma v))_F \ .$$

D'après le Théorème d'Hahn-Banach, $\tilde{f}$ peut être prolongé à F tout entier. Si $\tilde{\tilde{f}}$ est un prolongement de $\tilde{f}$, on peut alors définir $f_h \in V_h$ par

$$(f_h \ , \ v_h)_h = ((\tilde{\tilde{f}}, p_h \, v_h))_F \quad , \quad \forall v_h \in V_h \ . \qquad (1)$$

Cependant $\tilde{\tilde{f}}$ est en général impossible à déterminer.

---

(1) On rappelle que $(,)_h$ désigne la dualité sur V_h.

Le problème peut néanmoins être résolu lorsque $f \in H = L^2(\Omega)$.

Posons :

$$\tilde{\sigma}\, f = (f, 0, \ldots, 0) .$$

Alors $\tilde{\sigma}\, f \in F$ et

$$((\, \tilde{\sigma} f,\ p_h\, v_h))_F = (f,\ \pi\, p_h\, v_h)_H .$$

Mais

$$(f,\ \pi\, p_h\, v_h)_H = \sum_{M \in R_h \cap \Omega} \int_{\varpi_h^o(M) \cap \Omega} v_h^M \,.\, f(x)\ \theta_h^M(x)\, dx$$

et l'on peut poser

$$f_h(x) = \sum_{M \in R_h \cap \Omega} f_h^M\ \theta_h^M$$

avec

$$f_h^M = \int_{\varpi_h^o(M) \cap \Omega} f(x)\, dx \quad .$$

Dans ce cas

$$f_h \longrightarrow f \quad \text{dans} \quad H \quad \text{fort}$$

et les hypothèses (1.12) et (1.13) sont vérifiées. ∎

C H A P I T R E III

EXISTENCE ET UNICITE POUR LES PROBLEMES PARABOLIQUES

1. Formulation abstraite

Nous nous donnons :

(1.1) V, espace de Hilbert ; produit scalaire $((,))$; norme $\|\ .\ \|$

(1.2) V', dual de V ; produit scalaire canonique $((,))_*$; norme canonique $\|.\|_*$

(1.3) H, espace de Hilbert ; produit scalaire $(,)$; norme $|\ .\ |$

H', dual de H ; H' identifié à H ce qui permet de noter encore par $(,)$ le produit scalaire de la dualité entre V et V'

$V \subset H$ avec injection continue et densité : $V \hookrightarrow H \hookrightarrow V'$ (1)

(1.4) K, convexe fermé non vide de V

$u_o \in K$

(1.5) $t \to a(t;u,v)$, fonction mesurable sur $[0,T]$, $T < +\infty$, ceci $\forall u, v \in V$ avec pour presque tout $t \in [0,T]$,

$a(t;u,v)$, forme bilinéaire, monotone (ou positive) i.e.

$$a(t;u,v) \geqslant 0 \quad , \quad \forall\, v \in V \quad , \quad \text{p.p. en } t \in [0,T]$$

et continu au sens suivant

$$|a(t;u,v)| \leqslant C\, \|u\|.\|v\| \ , \quad \forall u, v \in V \ , \ \text{p.p. en } t \ ,$$

C constante indépendante de t,u,v ; ceci permet de définir un opérateur

$$A(t) : V \longrightarrow V' \quad \text{tel que} \quad (A(t)u,v) = a(t;u,v) \ , \ \forall\, u,v \in V$$

---

(1) Le symbole $\hookrightarrow$ désigne l'inclusion, la densité et l'injection continue.

(1.6) $t \longrightarrow j(t,v)$, fonction mesurable sur $[0,T]$, ceci $\forall v \in V$, avec

$|j(t,v)| \leqslant g(t)$, fonction intégrable indépendante de V , ceci $\forall v \in K$ [1] (ou bien l'hypothèse semblable :

$|j(t,v)| \leqslant C + C\|v\|$, C constante indépendante de t et v)

$j(t,v)$, fonctionnelle s.c.i. et convexe sur V , p.p. sur $[0,T]$

(1.7) $f \in L^2(0,T;V')$

(1.8) $W(0,T) = \{ v \mid v \in L^2(0,T;V) \ , \ v' \in L^2(0,T;V') \}$

(1.9) $W_o(0,T) = \{ v \mid v \in w(0,T) \ , \ v(0) = u_o \}$

(1.10) $\mathcal{K} = \{ v \mid v \in W_o(0,T) \ , \ v \in K \ , \ \text{p.p. sur } [0,T] \}$

(1.11) $\mathcal{K}_R = \{ v \mid v \in \mathcal{K}, \|v\|_{L^2(0,T;V)} \leqslant R_1 \ , \|v'\|_{L^2(0,T;V')} \leqslant R_2 \}$

avec $R = (R_1, R_2)$, R_1 , $R_2 \in \mathbb{R}^+ \cup \{+\infty\}$ [2]

(1.12) $\mathcal{K}_f = \{ v \mid v \in L^2(0,T;V) \ , \ v \in K \ , \ \text{p.p. sur } [0,T] \}$ [3].

Nous allons étudier les inéquations suivantes [4]

$$\int_0^T (u' + Au - f, v-u)dt + \int_0^T [j(v) - j(u)]\, dt \geqslant 0, \ \forall v \in \mathcal{K}_R \ , \ u \in \mathcal{K}_R ,$$

$$\int_0^T (v' + Av - f, v-u)dt + \int_0^T [j(v) - j(u)]\, dt \geqslant 0, \ \forall v \in \mathcal{K} \ , \ u \in \mathcal{K}_f .$$

2. Résultats préliminaires

Lemme 2.1. Muni du produit scalaire

$$((u,v))_{W(0,T)} = ((u,v))_{L^2(0,T;v)} + ((u',v'))_{L^2(0,T;V')}$$

l'espace $W(0,T)$ est un espace de Hilbert. De plus, si $v \in W(0,T)$, en modifiant éventuellement v sur un ensemble de mesure nulle, v est continu de $[0,T]$ dans H. [5] ■ Ce lemme rend licite la condition $v(0) = u_o$ dans la définition de $W_o(0,T)$. La deuxième partie de ce lemme s'établit à partir du

Lemme 2.2.

L'espace $\mathcal{C}^\infty([0,T] ; V)$ est dense dans $W(0,T)$ [6] . ■

La formule de Green étant justifiée pour les fonctions $\mathcal{C}^\infty([0,T] ; V)$ on a le

Lemme 2.3.

Si $u, v \in W(0,T)$, on a :

---

(1) Cette hypothèse est justifiée dans la suite (cf. Lemme 2.8) .

(2) $\|v\|_{L^2(0,T;V)} = \int_0^T \|v\|^2 \, dt$, $\|v\|_{L^2(0,T;V')} = \int_0^T \|v'\|^2 \, dt$.

(3) $\mathcal{K}_f$, convexe "faible" .

(4) Afin d'alléger les notations on omet d'écrire la variable t.

(5) cf. LIONS [1] , CEA-GOULAOUIC [1], RAVIART [1] [2] .

$$\int_0^t [(u',v) + (u,v')]\,dt = (u(t), v(t)) - (u(0), v(0))$$ (1) ∎

Lemme 2.4.

Si $v \in W(0,T)$, l'écriture suivante est justifiée :

$$\int_0^T (v',v)dt = \int_0^T \tfrac{1}{2}\frac{d}{dt}(v,v)dt = \tfrac{1}{2}|v(T)|^2 - \tfrac{1}{2}|v(0)|^2 .$$ (1) ∎

Lemme 2.5.

L'intégrale $\int_0^T (v',v)dt$ est une fonction convexe sur $W_o(0,T)$. ∎

Lemme 2.6.

L'intégrale $\int_0^T (v',v)dt$ est une fonction continue sur $W(0,T)$. ∎

La démonstration de ces deux derniers lemmes s'établit sans difficulté.

Lemme 2.7.

Soit $\varphi(t,v)$ une fonction numérique de $[0,T] \times V$ dans $\mathbb{R} \cup \{+\infty\}$, s.c.i. sur V pour presque tout $t \in [0,T]$, telle que $t \to \varphi(t,v(t))$ soit mesurable $\forall\ v \in L^1(0,T;V)$ et telle que $|\varphi(t,v(t))| \leqslant g(t)$ $\forall\ v \in L^1(0,T;V)$ avec $g(t)$ intégrable sur $[0,T]$. Alors la fonction

$$\Phi(v) = \int_0^T \varphi(t,v)dt$$

est s.c.i. sur $L^1(0,T;V)$. ∎

Démonstration

Nous allons montrer que $S(\lambda) = \{ v \mid v \in L^1(0,T;V) , \int_0^T \varphi(t,v)dt \leqslant \lambda \}$ est fermé $\forall\ \lambda \in \mathbb{R}$. Pour cela, nous considérons une suite v_ε , $v_\varepsilon \to v$ dans $L^1(0,T;V)$ et telle que $\int_0^T \varphi(t,v_\varepsilon)dt \leqslant \lambda$. On peut extraire (2) de v_ε une sous-suite $v_{\varepsilon'}$ qui converge presque partout vers v sur $[0,T]$. Alors, pour presque tout $t \in [0,T]$, on a

$$\liminf \varphi(t,v_{\varepsilon'}) \geqslant \varphi(t,v) .$$

D'après la condition $|\varphi(t,v(t))| \leqslant g(t)$, on peut appliquer le Théorème d'intégration terme à terme de Fatou-Lebesgue (3) d'où il résulte que $\liminf \varphi(t,v_{\varepsilon'})$ est intégrable et que

$$\int_0^T \liminf \varphi(t,v_{\varepsilon'})dt \leqslant \liminf \int_0^T \varphi(t,v_{\varepsilon'})dt \leqslant \lambda .$$

---

(1) cf. LIONS [1] , CEA-GOULAOUIC [1] , RAVIART [1] [2].

(2) YOSIDA [1] p.53.

(3) YOSIDA [1] p. 17 , ZAANEN [1] p. 48, BOURBAKI [1] , Intégration Ch. IV, § 3 , No. 7 , p. 140 .

Toujours d'après la condition $|\varphi(t,v(t))| \leqslant g(t)$, on a $\varphi(t,v)$ intégrable et utilisant un théorème classique [1] on déduit de $\varphi(t,v) \leqslant \liminf \varphi(t,v_{\varepsilon'})$ que

$$\int_0^T \varphi(t,v)dt \leqslant \int_0^T \liminf \varphi(t,v_{\varepsilon'})dt \leqslant \lambda .$$

Donc

$$\varphi(t,v) \in S(\lambda) \quad \text{et} \quad S(\lambda) \quad \text{est fermé.}$$

Le Lemme est valable sur $L^p(0,T;V)$, $1 \leqslant p \leqslant \infty$.

Lemme 2.8.

Avec les hypothèses (1.5) (1.6) les fonctionnelles :

$$v \longrightarrow \int_0^T (A(t)\,v,v)dt$$

et

$$v \longrightarrow \int_0^T j(t,v)dt$$

sont convexes sur $L^2(0,T;V)$. La première est continue et la seconde est s.c.i. sur $B(R) = \{ v \mid v \in L^2(0,T;V) , \|v\|_{L^2(0,T;V)} \leqslant R \}$ ceci $\forall R$, $R \leqslant \sup_{v \in K} \|v\| . \sqrt{T}$.

Démonstration

Comme $(A(t)v,v)$ et $j(t,v)$ sont mesurables et majorées en valeur absolue par des fonctions intégrales sur $[0,T]$, les intégrables sont bien définies. La convexité de la première fonctionnelle est une conséquence immédiate de la linéarité de l'intégrale et du Lemme 2.2. Ch. I . Pour la seconde cela se déduit de (1.6). La semi-continuité-inférieure est une conséquence du Lemme 2.7 [2] .

Lemme 2.9.

Si K n'est pas borné et si l'on a l'une des deux hypothèses $a(t;u,v)$ est coercif au sens suivant (nous dirons que $a(t;u,v)$ est t-coercif) :

$$a(t;v,v) \geqslant \alpha \|v\|^2 \quad , \forall\, v \in V \ , \alpha > 0$$

constante indépendante de t et de v , ceci p.p. sur $[0,T]$ ou bien $j(t,v)$ est fortement convexe au sens suivant (nous dirons que $j(t,v)$ est t- fortement convexe) :

1) $j(t,v)$ est convexe en $v \in K$.
2) $j\left(t, \frac{v_1 + v_2}{2}\right) - j(t,v_1) \leqslant \frac{1}{2}\left[j(t,v_2) - j(t,v_1)\right] - \beta \|v_2 - v_1\|^2$,

$\forall\, v_1, v_2 \in K$, avec $\beta > 0$ indépendant de t , de v_1 et v_2 , et ceci p.p. sur $[0,T]$.

---

(1) ZAANEN [1] p. 46.

(2) On pourra noter que la topologie de $L^2(0,T;V)$ est plus fine que celle de $L^1(0,T;V)$.

alors il est possible de définir un convexe fermé borné $\widetilde{K} \subset V$ tel que les deux problèmes suivants soient équivalents :

$$\int_0^T (v' + Au - f,\ v-u)dt + \int_0^T [j(v) - j(u)]\,dt \geqslant 0\ ,\quad \forall v \in \mathcal{K}_R\ ,\ u \in \mathcal{K}_{R'}\ ,$$

$$\int_0^T (\tilde{v}' + A\tilde{u} - f,\ \tilde{v}-\tilde{u})dt + \int_0^T [j(\tilde{v}) - j(\tilde{u})]\,dt \geqslant 0\ ,\quad \forall \tilde{v} \in \widetilde{\mathcal{K}}_R\ ,\ \tilde{u} \in \widetilde{\mathcal{K}}_{R'}\ ,$$

où $\widetilde{\mathcal{K}}_R$ et $\widetilde{\mathcal{K}}_{R'}$ sont définis comme $\mathcal{K}_R$ et $\mathcal{K}_{R'}$ en remplaçant K par $\widetilde{K}$,

$(R = (R_1\ ,\ R_2)\ ,\ R' = (R_1'\ ,\ R_2')\ ,\ R_1\ ,R_2\ ,\ R_1'\ ,\ R_2'\ ,\ \in \mathbb{R}^+ \cup \{+\infty\})$. ∎

Démonstration

La démonstration se fait comme pour le Lemme 2.4, Ch. I. On utilise cependant ici les résultats du Lemme 2.8. ∎

3. Existence dans le cas borné : solutions fortes

Nous donnons un premier théorème d'existence dans le cas de convexes bornés dans $W(0,T)$.

Théorème 3.1.

Sous les hypothèses (1.1)...(1.7) avec $R = (R_1\ ,\ R_2)\ ,\ R' = (R_1'\ ,\ R_2')$ $R_1\ ,\ R_2\ ,\ R_1'\ ,\ R_2' \in \mathbb{R}^+(\neq +\infty)\ ,\ R_1' \geqslant R_1\ ,\ R_2' \geqslant R'$, les problèmes :

$$(3.1)\qquad \int_0^T (u' + Au - f,\ v-u)dt + \int_0^T [j(v) - j(u)]\,dt \geqslant 0\ ,\quad \forall v \in \mathcal{K}_R\ ,\ u \in \mathcal{K}_{R'}\ ,$$

$$(3.2)\qquad \int_0^T (v' + Av - f,\ v-u)dt + \int_0^T [j(v) - j(v)]\,dt \geqslant 0\ ,\quad \forall v \in \mathcal{K}_R\ ,\ u \in \mathcal{K}_{R'}\ ,$$

ont chacun au moins une solution. En outre, si $R_1 = R_1'$, $R_2 = R_2'$ les problèmes (3.1)(3.2) sont équivalents. ∎

Démonstration

On introduit

$$L(u,v) = \int_0^T (u' + Au - f\ ,\ v-u)dt + \int_0^T [j(v) - j(u)]\,dt$$

et l'on vérifie sans difficulté les conditions du Théorème 2.1., Ch. I, en sorte qu'il existe au moins un point selle $(u_1\ ,\ u_2) \in \mathcal{K}_R \times \mathcal{K}_{R'}$. On vérifie alors aisément que cela assure l'existence de solutions pour (3.1) et (3.2). L'équivalence n'offre pas de difficulté. ∎

4. Existence dans le cas borné : solutions faibles

Nous allons affaiblir les hypothèses précédentes et démontrer le

Théorème 4.1.

Sous les hypothèses (1.1)...(1.7) avec $R = (R_1 , R_2) , R_1 , R_2 \neq +\infty$, et K borné ou a(t;u,v) t-coercif ou j(t,v) t-fortement convexe, le problème

$$(4.1) \qquad \int_0^T (v' + Av - f , v-u)dt + \int_0^T [j(v) - j(u)]\, dt \geqslant 0 \quad , \forall\, v \in \mathcal{K}_R , u \in \mathcal{K}_f$$

a au moins une solution. ∎

Démonstration

On introduit

$$L(u,v) = \int_0^T (v' + Av - f, v-u)dt + \int_0^T [j(v) - j(u)]\, dt$$

et l'on vérifie sans difficulté que les conditions sur L(u,v) requises dans le Théorème 2.1., ch. I, sont assurées. Les hypothèses portant sur les convexes sont immédiates : K borné $\overset{(1)}{\Longrightarrow}$ $\mathcal{K}_f$ borné dans $L^2(0,T;V)$, et $\mathcal{K}_R$ borné par hypothèse dans W(0,T). on peut donc appliquer le Théorème de Ky-Fan-Sion en sorte qu'il existe $(u_1 , u_2) \in \mathcal{K}_f \times \mathcal{K}_R$ tel que

$$L(v,u_2) \leqslant L(u_1,u_2) \leqslant L(u_1 , w) , \quad \forall\, (v,w) \in \mathcal{K}_f \times \mathcal{K}_R .$$

Or $\mathcal{K}_f \supset \mathcal{K}_R$, on peut donc prendre $v = u_2$, d'où

$$L(u_1 , w) \geqslant 0 \quad , \quad \forall\, w \in \mathcal{K}_R \quad , \quad u_1 \in \mathcal{K}_f$$

et ceci prouve l'existence de u_1 solution de (4.1). ∎

Théorème 4.2.

Sous les hypothèses (1.1)...(1.7) avec K borné ou a(t;u,v) t-coercif ou j(t,v) t-fortement convexe, le problème

$$(4.2) \qquad \int_0^T (v' + Av - f, v-u)dt + \int_0^T [j(v) - j(u)]\, dt \geqslant 0 \ , \forall\, v \in \mathcal{K} , \ u \in \mathcal{K}_f$$

a au moins une solution. ∎

---

(1) Dans les démonstrations nous n'analysons pas les hypothèses a(t;u,v) t-coercif et j(t,v) t-fortement convexe. Nous avons vu en effet que ces deux dernières hypothèses équivalaient à considérer le convexe K comme borné.

Démonstration

Notons u_R une des solutions $u \in \mathcal{K}_f$ qui vérifie (4.1) ($\forall\, v \in \mathcal{K}_R$). Lorsque $R \to +\infty$ (R_1, $R_2 \to +\infty$), la suite u_R reste dans un convexe fermé borné $\mathcal{K}_f$ de $L^2(0,T;V)$. On peut donc extraire une sous-suite qui tend faiblement vers un élément $u \in \mathcal{K}_f$ et en passant à la limite dans (4.1) on obtient (4.2). ∎

Remarque 4.1.

Nous avons considéré

$$\mathcal{K}_R = \{ v \mid v \in \mathcal{K} \ , \ \|v\|_{L^2(0,T;V)} \leq R_1 \ , \ \|v'\|_{L^2(0,T;V')} \leq R_2 \}$$

mais l'hypothèse $\|v\|_{L^2(0,T;V)} \leq R_1$ est inutile dès que l'hypothèse

$$\|v'\|_{L^2(0,T;V')} \leq R_2 < +\infty \qquad \text{est}$$

vérifiée. En effet, en utilisant la densité de $\mathcal{D}([0,T] ; V)$ dans $W(0,T)$ et ayant

$$\varphi(t) = \int_0^T \varphi'(s)ds + \varphi(0) \ , \quad \forall\, \varphi \in \mathcal{D}([0,T] ; V)$$

on en déduit aisément que

$$\|v\|_{L^2(0,T;V)} \leq R_2\, T + \text{constante}$$

donc v est borné dans $L^2(0,T;V)$. ∎

5. Problèmes pseudo-instantanés équivalents

Théorème 5.1.

Les problèmes (3.1) et

$$(5.1) \qquad (u' + Au - f \ , \ v-u) + j(v) - j(u) \geq 0 \ , \forall\, v \in \mathcal{K}_R \ , \ u \in \mathcal{K}_R \ .$$

avec :

$$R = (R_1 \ , \ R_2) \ , \ R_1 \ , \ R_2 \in \mathbb{R} \cup \{+\infty\}$$

et ceci $\forall\, t \in [0,T]$, sauf peut-être pour un ensemble $Z \subset [0,T]$ de mesure nulle, sont équivalents. ∎

Démonstration

On vérifie facilement que l'on peut intégrer (5.1) sur $[0,T]$ et l'on obtient (3.1). Réciproquement, il faut montrer que si u est solution de (3.1) alors u est solution de (5.1).

Soient $w \in \mathcal{K}_R$ et $\varphi(t) \in \mathcal{D}(0,T)$, $\varphi(t) \geqslant 0$. Posons, si $\varphi(t) \not\equiv 0$,

$\psi(t) = \varphi(t) / \max_{t\in[0,T]} |\varphi(t)|$, sinon $\psi(t) \equiv 0$. Alors $\psi(t) \in [0,1]$, $\forall t \in [0,T]$.

Soit

$$v(t) = u(t) + \psi(t)\,[w(t) - u(t)]\,.$$

Alors $v(t) \in \mathcal{K}$ et (3.1) s'écrit

$$\int_0^T (u' + Au - f,\ w-u)\psi(t)dt + \int_0^T [j(u + \psi(w-u)) - j(u)]\,dt \geqslant 0$$

et d'après la convexité

$$\int_0^T (u' + Au - f,\ w-u)\psi(t)dt + \int_0^T [j(w) - j(u)]\ \psi(t)dt \geqslant 0\ .$$

Comme ceci est vrai $\forall\ \psi(t) \in \mathcal{D}(0,T)$ et $0 \leqslant \psi(t) \leqslant 1$ on a

$$\int_0^T (u' + Au - f,\ w-u)\varphi(t)dt + \int_0^T [\,j(w) - j(u)]\varphi(t)dt \geqslant 0\,,$$

ceci $\forall\ \varphi(t) \geqslant 0$, $\varphi(t) \in \mathcal{D}(0,T)$. Autrement dit, sauf peut-être pour $t \in Z \subset [0,T]$ Z de mesure nulle, on a

$$(u' + Au - f,\ w-u) + j(w) - j(u) \geqslant 0\ ,\ \forall\ w \in \mathcal{K}_R\ ,$$

ce qui prouve que les inéquations (3.1) et (5.1) sont équivalentes. ∎

Remarque 5.1.

Pour le problème (4.2) et son inéquation pseudo-instantanée :

$$(5.2) \qquad (v' + Av - f,\ v-u) + j(v) - j(u) \geqslant 0\ ,\ \forall v \in \mathcal{K}\ ,\ u \in \mathcal{K}_f$$

on ne peut pas poser directement $v = u + \psi\,.\,[w-u]$ car alors v peut ne plus appartenir à $L^2(0,T;V')$ $(u \in \mathcal{K}_f)$. On a cependant toujours (5.2) $\Rightarrow$ (4.2). ∎

6. Problèmes instantanés équivalents

Dans le paragraphe précédent les problèmes étaient considérés avec $v \in \mathcal{K}$. Ici nous affaiblissons cette condition.

Théorème 6.1.

Les problèmes :

$$(6.1) \qquad (u' + Au - f,\ v-u) + j(v) - j(u) \geqslant 0\ ,\ \forall v \in K\ ,\ u \in \mathcal{K}_R\ ,$$

et ceci $\forall t \in [0,T]$, sauf peut-être pour un ensemble $Z \subset [0,T]$ de mesure nulle

(6.2) $$\int_0^T (u' + Au - f,\ v-u)dt + \int_0^T [j(v) - j(u)]\,dt \geqslant 0 \quad , \ \forall v \in \mathcal{K} \ , u \in \mathcal{K}_R$$

avec

$$R = (R_1 , R_2),\ R_1, R_2 \in \mathbb{R}^+ \cup \{+\infty\} \ , \ \underline{\text{sont équivalents}}. \quad \blacksquare$$

Démonstration

Montrons que u solution de (6.2) est aussi solution de (6.1). Soient $w \in K$ et ψ comme précédemment. Soit $v(t) = u(t) + \psi(t).[w-u(t)]$. Alors $v(t) \in \mathcal{K}$. Par définition de u , on a

$$\int_0^T (u' + Au - f,\ v(t)-u(t))dt + \int_0^T [j(v(t)) - j(u(t))]\,dt \geqslant 0 \quad .$$

Donc, avec la convexité de $j(v)$,

$$\int_0^T (u' + Au - f,\ w-u)\ \psi(t)dt + \int_0^T [j(w) - j(u)]\ \psi(t) \geqslant 0$$

et l'on obtient bien (6.1).

Réciproquement, soit u solution de (6.1). On considère une subdivision de $[0,T]$ en $N(k)$ intervalles égaux de longueur k : $[0,k,2k,\ldots, N(k).k = T]$

Nous faisons l'hypothèse suivante (cf. Ch. IV).

$$\begin{cases} \forall\, w \in \mathcal{K} \ , \exists\ w_k(t) = \sum_{i=1}^{N(k)} \psi_{k,i}(t) v_i \quad \text{avec} \\ v_i \in K \ , \quad \psi_{k,i}(t) \in [0,1] \ , \quad \forall\, t \ , \quad \psi_{k,i} \in L^1(0,T), \\ \sum_{i=1}^{N(k)} \psi_{k,i}(t) = 1 \ , \text{ p.p. sur } [0,T] \\ \text{tel que} \\ \quad \left| \begin{array}{l} w_k \longrightarrow w \text{ dans } \mathcal{K} \text{ fort} \ , \\ \limsup \int_0^T \sum_{i=1}^{N(k)} \psi_{k,i}(t)\ j(v_i)dt \leqslant \int_0^T j(w)dt \ . \end{array} \right. \end{cases}$$

En prenant $v = v_i$ dans (6.1), en multipliant par $\psi_{k,i}$ et en intégrant, il vient

$$\int_0^T (u' + Au - f,\ v_i - u)\ \psi_{k,i}(t)dt + \int_0^T [j(v_i) - j(u)]\psi_{k,i}(t)dt \geqslant$$

En sommant pour $i=1,\ldots, N(k)$, en utilisant l'hypothèse ci-dessus et en passant à la limite on obtient (6.2). $\blacksquare$

7. Unicité

Théorème 7.1.

Si $a(t;u,v)$ est t-coercif et /ou si $j(t,v)$ est t-fortement convexe le problème (3.1) (avec $R \neq +\infty$ uniquement pour assurer l'existence) a une solution unique. Avec les mêmes hypothèses et $j(t,v)$ continu sur $L^2(0,T;V)$ les problèmes (4.1) (4.2) ont des solutions uniques. ∎

Démonstration

Pour (3.1) on obtient, pour deux solutions éventuelles $u_1 \neq u_2$:

$$\tfrac{1}{2}\,|u_1(T) - u_2(T)|^2 + \int_0^T \|u_1 - u_2\|^2\,dt \leqslant 0$$

d'où l'unicité.

Pour (4.2) par exemple (cf. LIONS [1][5]) on pose $v = \frac{u_1 + u_2}{2}$ et l'on approche v par une suite v_ε telle que $\varepsilon v_\varepsilon' + v_\varepsilon = v$. Alors $v_\varepsilon \in K$ et prenant $v = v_\varepsilon$ dans (4.2) écrit respectivement pour u_1 et u_2 , en ajoutant, et en remarquant que :

$$2\varepsilon \int_0^T (v_\varepsilon' , v_\varepsilon - v)dt = -2\varepsilon \int_0^T |v_\varepsilon'|^2 dt \leqslant 0 ,$$

$$v_\varepsilon \longrightarrow v \quad \text{dans} \quad L^2(0,T;V) ,$$

on obtient (avec $j(v)$ continu)

$$\int_0^T [(Au_1, v-u_1) + (Au_2, v-u_2) + 2\,j(v) - j(u_1) - j(u_2)]\,dt \geqslant 0 ,$$

d'où

$$-\tfrac{1}{2}\int_0^T (Au_1 - Au_2 , u_1 - u_2)dt + 2\int_0^T [\,j(\tfrac{u_1 + u_2}{2}) - j(u_1) - j(u_2)]\,dt \geqslant 0$$

et l'unicité s'en déduit immédiatement. ∎

C H A P I T R E IV

APPROXIMATION DES PROBLEMES PARABOLIQUES

Nous allons traiter dans ce chapitre de l'approximation externe des problèmes paraboliques.

1. Théorie

Orientation

Nous considérons les problèmes :

$$(1.1)\qquad \int_0^T (u' + Au - f,\ v-u)dt + \int_0^T [j(v) - j(u)]\,dt \geqslant 0, \quad \forall v \in K ,\ u \in K ,$$

$$(1.2)\qquad \int_0^T (v' + Av - f,\ v-u)dt + \int_0^T [j(v) - j(u)]\,dt \geqslant 0, \quad \forall v \in K ,\ u \in K_f .$$

Rappelons que (1.1) est équivalent à [1]

$$(1.3)\qquad (u' + Au - f\ ,\ v-u) + j(v) - j(u) \geqslant 0 \ ,\ \forall v \in K \ ,\ u \in K$$

et que (1.2) est résolu par

$$(1.4)\qquad (v' + Av - f,\ v-u) + j(v) - j(u) \geqslant 0 \ ,\ \forall v \in K \ ,\ u \in K_f .$$

Nous allons définir une famille d'inéquations "proches" de (1.3) et dont il est possible de calculer "plus simplement" les solutions. Les solutions de ces

---

(1) Pour les problèmes instantanés considérés par la suite il est sous-entendu que les inéquations sont définies , $\forall t \in [0,T]$ sauf peut-être sur un sous-ensemble $Z \subset [0,T]$ de mesure nulle.

inéquations seront cherchées dans des espaces $L^2(0,T;V_h)$, V_h de dimension finie. Appelant u_h l'une de ces solutions nous allons montrer que :

$$"u_h" \longrightarrow u \quad \text{dans } L^2(0,T;V) \text{ faible,}$$

u étant solution "faible" de l'inéquation (1.2), et dans le cas où (1.1) possède une solution forte unique u nous montrerons que :

$$"u_h" \longrightarrow u \quad \text{dans } L^2(0,T;V) \text{ fort. } \blacksquare$$

1.1. Cadre fonctionnel.

Nous conservons les notations, espaces et opérateurs définis au § 1.1 du Ch. II. Nous introduisons cependant un second produit scalaire et une seconde norme sur V_h :

$$(v_h , w_h)_h \quad \text{et} \quad |v_h|_h$$

qui "approximent" le produit scalaire $(,)$ et la norme $|.|$ sur H. On définit un prolongement q_h de V_h dans H, linéaire et injectif tel que $q_h = \pi\, p_h$ et l'on définit $(,)_h$ et $|.|_h$ par

$$(1.4) \qquad (v_h , w_h)_h = (q_h v_h , q_h w_h) ,$$

$$(1.5) \qquad |v_h|_h = |q_h v_h| .$$

On suppose que

$$(1.6) \qquad |v_h|_h \leqslant C \, \|v_h\|_h , \quad (\text{de même que } |.| \leqslant C\|.\|) ,$$

$$(1.7) \qquad \|v_h\|_h \leqslant S(h)|v_h|_h , \quad S(h) \longrightarrow +\infty .$$

On se donne une subdivision de $[0,T]$ en $N = N(k)$ intervalles de longueur k, k étant destiné à tendre vers 0, correlativement $N(k) \longrightarrow +\infty$. Nous cherchons une "approximation" de u(t) dans la classe des fonctions en escalier sur $[0,T]$ à valeurs dans V_h .
Soient $u_h^i \in V_h$, $i = 0,\ldots, N$.
Soit $\chi_k^i(t)$ la fonction caractéristique de $[ik, (i+1)k[$.
On pose

$$u_{h,k}(t) = \sum_{i=0}^{N} u_h^i \, \chi_k^i(t) \quad (1) .$$

---

(1) éventuellement on définira $u_{h,k}(t)$ sur $[0,T+k]$ en sommant de 0 à N+1 .

On définit la "dérivée discrète" de $u_{h,k}(t)$ par

$$\delta_k u_{h,k}(t) = \frac{u_{h,k}(t+k) - u_{h,k}(t)}{k}$$

ou par

$$\bar{\delta}_k u_{h,k}(t) = \frac{u_{h,k}(t) - u_{h,k}(t-k)}{k} = \delta_k u_{h,k}(t-k) \quad . \quad \blacksquare$$

1.2. Approximation de $\mathcal{K}$ (ou de $W_o(0,T)$ lorsque $\mathcal{K} = W_o(0,T)$)

On reprend (1.4) (1.10) (1.12) Ch. III pour la définition de $K, \mathcal{K}, \mathcal{K}_f$.
On se donne

$$K_h \text{ convexe fermé de } V_h \quad .$$

On supposera uniquement pour simplifier les démonstrations que

$$0 \in K \quad , \quad 0 \in K_h \quad .$$

On suppose que l'approximation est consistante :

(1.10)

$v_{h,k}(t) \in K_h$, $t \in [0,T]$,

$q_h v_{h,k} \longrightarrow v$ dans $L^\infty(0,T;H)$ faible étoile ,

$p_h v_{h,k} \longrightarrow \tilde{v}$ dans $L^2(0,T;F)$ faible

$\Rightarrow \quad v \in K$, p.p. sur $[0,T]$ et $\tilde{v} = \sigma v$,

et qu'elle est convergente :

(1.11)

$\forall v \in L^2(0,T;K)$, $\exists v_{h,k} \in L^2(0,T;K_h)$ (h n'étant pas lié à k

tel que :

$q_h v_{h,k} \longrightarrow v$ dans $L^2(0,T;H)$ fort ,

$p_h v_{h,k} \longrightarrow \sigma v$ dans $L^2(0,T;F)$ fort

(1.12.)

et si de plus $v' \in L^2(0,T;V')$ alors

$q_h \quad \delta_k v_{h,k} \longrightarrow v'$ dans $L^2(0,T;V')$ fort .

(1.13) Par ailleurs on se donne $u_o \in K$ et l'on suppose qu'il existe $u_{o,h} \in K_h$ tel que

$$\left| \begin{array}{l} p_h\, u_{o,h} \longrightarrow \sigma u_o \quad \text{dans } F \text{ fort}, \\ \|u_{o,h}\|_h \leqslant C \qquad (h \to 0). \end{array} \right. \quad \blacksquare$$

1.3. Approximation de $a(t;u,v)$

On reprend les hypothèses faites sur $a(u,v)$ ((1.5) Ch. III). Pour chaque $t \in [0,T]$ on se donne une famille de formes bilinéaires sur $L^2(0,T;V_h)$:

(1.14)
$$\left| \begin{array}{l} t \longrightarrow a_{h,k}(t;u_{h,k}, v_{h,k}) \text{ mesurable sur } [0,T] \\ \text{constante en } t \text{ sur chaque intervalle}, \end{array} \right.$$

continues au sens suivant :

(1.15)
$$\left| \begin{array}{l} |a_{h,k}(t;u_h, v_h)| \leqslant C \|u_h\|_h \|v_h\|_h \quad , \forall u_h \in V_h \quad , \forall v_h \in V_h \\ C \text{ constante indépendante de } h, \text{ de } k, \text{ et de } t, \end{array} \right.$$

coercives au sens suivant :

(1.16)
$$\left| \begin{array}{l} a_{h,k}(t;u_h, u_h) \geqslant \alpha \|u_h\|_h^2 \quad , \quad \forall u_h \in V_h, \\ \alpha \text{ constante} > 0 \text{ indépendante de } h, \text{ de } k, \text{ et de } t, \end{array} \right.$$

consistantes, i.e.

(1.17)
$$\left| \begin{array}{l} v_{h,k}(t) \in K_h \quad , \quad \forall\, t \in [0,T], \\ p_h\, v_{h,k} \longrightarrow \sigma v \text{ dans } L^2(0,T;F) \text{ faible}, \\ \Rightarrow \liminf \int_0^T a_{h,k}(t;v_{h,k}, v_{h,k})dt \geqslant \int_0^T a(t;v,v)dt, \end{array} \right.$$

convergentes, i.e.

(1.18)
$$\forall v \in L^2(0,T;K), \quad \forall w \in L^2(0,T;K),$$
si $v_{h,k} \in L^2(0,T;K_h)$ et $w_{h,k} \in L^2(0,T;K_h)$ vérifient
$$p_h\, v_{h,k} \longrightarrow \sigma v \text{ dans } L^2(0,T;F) \text{ faible},$$
$$p_h\, w_{h,k} \longrightarrow \sigma w \text{ dans } L^2(0,T;F) \text{ fort},$$
alors
$$\int_0^T a_{h,k}(t;v_{h,k}, w_{h,k})dt \to \int_0^T a(t;v,w)dt ,$$
$$\int_0^T a_{h,k}(t;w_{h,k}, v_{h,k})dt \to \int_0^T a(t;w,v)dt .$$

On définit $A_{h,k}(t) \in \mathcal{L}(V_h , V_h)$ par

(1.19)
$$(A_{h,k}(t)\, u_h, v_h)_h = a_{h,k}(t;u_h,v_h) , \quad \forall u_h , v_h \in V_h .$$

De (1.15) et (1.16) on déduit immédiatement que $A_{h,k}(t)$ est continu et coercif sur V_h. ∎

1.4. Approximation de $j(t,v)$

On reprend les hypothèses faites sur $j(t,v)$ ((1.6) Ch. III) . Pour chaque $t \in [0,T]$ on se donne une famille de fonctionnelles convexes sur $L^2(0,T;V_h)$:

(1.20)
$$t \longrightarrow j_{h,k}(t,u_{h,k}) \text{ mesurable sur } [0,T] ,$$
constante en t sur chaque intervalle ,

(1.21)
$$|j_{h,k}(t,u_{h,k})| \leqslant g_{h,k}(t) \text{ intégrable}, \quad \forall u_{h,k} \in L^2(0,T;K_h) ,$$

semi-continues inférieurement au sens suivant :

(1.22)
$$v_{h,k} \longrightarrow u_{h,k} \text{ dans } L^2(0,T;K_h) \text{ fort}$$
$$\Rightarrow \liminf \int_0^T j_{h,k}(t,v_{h,k})dt \geqslant \int_0^T j_{h,k}(t,u_{h,k})dt ,$$

consistantes, i.e.

(1.23)
$$v_{h,k}(t) \in K_h , \quad \forall t \in [0,T] ,$$
$$p_h\, v_{h,k} \longrightarrow \sigma v \text{ dans } L^2(0,T;F) \text{ faible}$$
$$\Rightarrow \liminf \int_0^T j_{h,k}(t,v_{h,k})dt \geqslant \int_0^T j(t,v)dt ,$$

convergentes, i.e.

(1.24)
$$v_{h,k}(t) \in K_h \ , \quad \forall t \in [0,T] \ ,$$
$$p_h v_{h,k} \longrightarrow \sigma v \text{ dans } L^2(0,T;F) \text{ fort}$$
$$\Rightarrow \int_0^T j_{h,k}(t,v_{h,k})dt \to \int_0^T j(t,v)dt \ . \quad \blacksquare$$

1.5. Approximation de $f(t)$

On se donne $f \in L^2(0,T;V')$ et une famille de fonctions $f_{h,k} \in L^2(0,T;V_h)$

(1.25)
$$t \to f_{h,k}(t) \text{ mesurable sur } [0,T] \ ,$$
constante en t sur chaque intervalle,

(1.26)
$$\|f_{h,k}(t)\|_h^* \leqslant C \ , \text{ constante indépendante de } h \text{ et } k,$$

consistante et convergente :

(1.27)
$$v_{h,k}(t) \in K_h \ ,$$
$$p_h v_{h,k} \longrightarrow v \text{ dans } L^2(0,T;F) \text{ faible}$$
$$\Rightarrow \int_0^T (f_{h,k} \ , \ v_{h,k})_h dt \to \int_0^T (f,v)dt \quad \blacksquare$$

1.6. Schémas de discrétisation

Pour la commodité de l'écriture nous omettons de considérer que

$$(A(t)\,u,v) = a(t;u,v) \ , \ j(t,v) \ , \ f(t)$$

dépendent du temps. On note A_h , j_h , f_h les éléments discrets correspondants. Nous étudierons les schémas suivants :

Schéma I.

(Schéma non découplé totalement implicite)

(1.28)
On cherche $u_{h,k} = \sum_{i=0}^{N} u_h^i \ \chi_k^i$ tel que

$$u_h^o = u_{o,h}$$
$$(\bar{\delta}_k u_{h,k} + A_h u_{h,k} - f_h \ , \ v_h - u_{h,k})_h + j_h(v_h) - j_h(u_{h,k}) \geqslant 0 \ ,$$
$$\forall v_h \in K_h \ , \ u_{h,k}(t) \in K_h \ , \quad \forall t \in [0,T] \ .$$

De façon "ponctuelle" cette inéquation s'écrit :

(1.29)
Trouver u_h^i , $i=0,1,\dots,N$ tels que :

$$u_h^o = u_{o,h} \quad ,$$

$$\left(\frac{u_h^{i+1} - u_h^i}{k} + A_h u_h^{i+1} - f_h \, , \, v_h - u_h^{i+1} \right)_h + j_h(v_h) - j_h(u_h^{i+1}) \geqslant 0 \quad ,$$

$$\forall v_h \in K_h \, , \; u_h^{i+1} \in K_h \, .$$

Schéma II.

(Schéma non découplé , cas général avec $j=0$)

(1.30)
Trouver u_h^i , $i=0,1,\dots,N$ tels que :

$$u_h^o = u_{o,h} \, ,$$

$$\left(\frac{u_h^{i+1} - u_h^i}{k} + A_h u_h^{i+\theta} - f_h \, , \, v_h - u_h^{i+1} \right) \geqslant 0 \, , \; \forall v_h \in K_h \, , \; u_h^{i+1} \in K_h \, ,$$

avec

$$u_h^{i+\theta} = u_h^i + \theta \, (u_h^{i+1} - u_h^i) \, , \; \theta \in [0,1] \, .$$

Schéma III.

(Schéma découplé avec $j=0$). Trouver :

(1.31)
$$u_{h,k}^* = \sum_{i=0}^{N} u_h^{i+\frac{1}{2}} \, \chi_k^i \in L^2(0,T;V_h) \, ,$$

et

$$u_{h,k} = \sum_{i=0}^{N} u_h^i \, \chi_k^i \in L^2(0,T;K_h) \, ,$$

tels que :

$$u_h^o = u_{o,h} \quad ,$$

$$\frac{u_h^{i+\frac{1}{2}} - u_h^i}{k} + A_h(\theta^- u_h^i + \theta^* u_h^{i+\frac{1}{2}} + \theta^+ u_h^{i+1}) - f = 0 \quad ,$$

$$u_h^{i+1} = P_{K_h} \left[u_h^{i+\frac{1}{2}} \right] \quad \text{(projecteur au sens de la norme } (,)_h) \, ,$$

$$\theta^- \, , \, \theta^* \, , \, \theta^+ \geqslant 0 \, , \quad \theta^- + \theta^* + \theta^+ = 1 \, .$$

Remarque 1.1.

Les schémas non découplés sont un cas particulier des schémas découplés (faire $\theta^* = 0$ et l'on verra que si $u_{h,k}$ vérifie (1.31), il vérifie (1.30)). Nous les séparons afin de distinguer 2 techniques de démonstration. ∎

1.7. Lemmes préliminaires

Lemme 1.1.

Soit

$$u_{h,k} = \sum_{i=0}^{N} u_h^i \chi_h^i \quad \text{tel que}$$

$$\alpha \sum_{i=1}^{N} k \|u_h^i\|_h^2 \leq C + \sum_{i=1}^{N} k \left[j_{h,k}(0) - j_{h,k}(u_h^i) \right] .$$

Alors

$$\int_k^T \|u_{h,k}\|_h^2 \, dt \leq C \quad \text{indépendante de } h \text{ et } k$$

autrement dit $u_{h,k}$ est borné dans $L^2(0,T;F)$. ∎

Démonstration

La technique de démonstration est la même que celle de la Proposition 1.1. du Ch. II. Nous reprenons les points principaux. Tout d'abord si $\int_k^T \|u_{h,k}\|_h^2 dt \leq 1$, $\forall h,k$ le résultat est trivial. Supposons alors qu'il existe une suite (h,k), $h \to 0$, $k \to 0$ telle que

$$\int_k^T \|u_{h,k}\|_h^2 \, dt \geq 1 \quad \text{et} \quad \int_k^T \|u_{h,k}\|^2 \, dt \to +\infty .$$

Nous allons montrer que

$$\frac{\sum_{i=1}^{N} k \left[j_{h,k}(0) - j_{h,k}(u_{h,k}) \right]}{\int_0^T \|u_{h,k}\|_h^2 \, dt}$$

reste borné. Pour cela on utilise d'abord la convexité des $j_{h,k}(t,u_h)$ et l'on pose (cf. Proposition 1.1., Ch. II.) $\theta = \dfrac{1}{\int_0^T \|u_{h,k}\|_h^2 dt}$.

Le reste de la démonstration se fait pratiquement comme à la Proposition 1.1.,Ch.II. On utilisera les hypothèses (1.22) et (1.23) . ∎

Corollaire 1.1. Si

$$\alpha \sum_{i=1}^{N} k \|u_h^i\|_h^2 \leq C + \sum_{i=1}^{N} Ck \|u_h^i\|_h + \sum_{i=1}^{N} k \left[j_{h,k}(0) - j_{h,k}(u_h^i) \right]$$

alors

$$\int_k^T \|u_{h,k}\|_h^2 \, dt \leq C \text{ , constante indépendante de } h \text{ et } k . \quad ∎$$

Démonstration

Immédiate en se reportant au lemme précédent et à la Démonstration de la Proposition 1.1., Ch. II. ∎

Lemme 1.2.

Soient

$$v_{h,k} = \sum_{i=0}^{N} v_h^i \chi_k^i \text{ , } u_{h,k} = \sum_{i=0}^{N} u_h^i \chi_k^i ,$$

alors

$$(1.32)\qquad \left(\frac{v_h^{i+1}-v_h^i}{k},\ v_h^{i+1}-u_h^{i+1}\right)_h = \left(\frac{u_h^{i+1}-u_h^i}{k},\ v_h^{i+1}-u_h^{i+1}\right)_h + \frac{1}{2k}\left|v_h^{i+1}-u_k^{i+1}\right|_h^2 -$$

$$-\frac{1}{2k}\left|v_h^i-u_h^i\right|_h^2+\frac{1}{2k}\left|v_h^{i+1}-v_h^i-(u_h^{i+1}-u_h^i)\right|_h^2$$

et si l'on a

$$(1.33)\qquad \int_k^T (\delta_k u_{h,k}+A_h u_{h,k}-f_h;\ v_{h,k}-u_{h,k})_h dt+\int_k^T \left[j_h(v_{h,k})-j_h(u_{h,k})\right]dt \geqslant$$

on a aussi

$$(1.34)\qquad \int_k^T (\delta_k v_{h,k}+A_h u_{h,k}-f_h,\ v_{h,k}-u_{h,k})_h dt+\int_k^T \left[j_h(v_{h,k})-j_h(u_{h,k})\right]dt \geqslant$$

$$\geqslant \tfrac{1}{2}\left|v_{h,k}(T)-u_{h,k}(T)\right|_h^2-\tfrac{1}{2}\left|v_{h,k}(0)-u_{h,k}(0)\right|. \quad \blacksquare$$

Démonstration

Pour obtenir (1.32) on note que

$$(v_h^{i+1}-v_h^i,\ v_h^{i+1}-u_h^{i+1})_h=(u_h^{i+1}-u_h^i,\ v_h^{i+1}-u_h^{i+1})_h+(v_h^{i+1}-v_h^i-(u_h^{i+1}-u_h^i),\ v_h^{i+1}-u_h^{i+1})_h$$

et l'on utilise l'égalité

$$(a-b,\ a)=\tfrac{1}{2}|a|^2-\tfrac{1}{2}|b|^2+\tfrac{1}{2}|a-b|^2 .$$

On obtient (1.32). En sommant (1.32) de $i=0$ à $N-1$ on obtient

$$\int_k^T (\bar{\delta}_k u_{h,k},\ v_{h,k}-u_{h,k})_h dt \leqslant \int_k^T (\bar{\delta}_k v_{h,k},\ v_{h,k}-u_{h,k})_h dt -$$

$$-\tfrac{1}{2}\sum_{i=1}^{N}\left|v_h^i-u_h^i\right|_h^2+\tfrac{1}{2}\sum_{i=0}^{N-1}\left|v_h^i-u_h^i\right|_h$$

d'où l'on déduit (1.34) en reportant cette majoration dans (1.33). ∎

Remarque 1.2.

Ce lemme reste vrai en remplaçant $A_h u_{h,k}$ dans (1.33) et (1.34) par n'importe quelle expression, par exemple par $A_h \tilde{u}_{h,k}$ où $\tilde{u}_{h,k}$ est une fonction quelconque.

Si l'on a (1.33) avec $A_h u_{h,k}$ on obtient aussi l'inégalité (1.34) dans laquelle on remplace $A_h u_{h,k}$ par $A_h v_{h,k}$ (ceci résulte de la coercivité). ∎

1.8. Existence et stabilité pour le schéma I

Notons d'abord que l'existence de $u_{h,k}$ est une conséquence immédiate de la coercivité de l'opérateur $(I_h + A_h)$, I_h opérateur identité de V_h dans V_h .

Nous allons chercher des majorations convenables de $u_{h,k}$.

Avec $v = 0$ dans (1.29) on obtient

$$(u_h^{i+1} - u_h^i , u_h^i)_h + k(A_h u_h^{i+1} , u_h^{i+1})_h \leqslant k(f_h , u_h^{i+1})_h + k\left[j_h(0) - j_h(u_h^{i+1})\right] .$$

Or

$$2(a - b, a) = |a|^2 - |b|^2 + |a - b|^2 \tag{1.35}$$

et

$$2(a,b) \leqslant \frac{1}{\varepsilon} |a|^2 + \varepsilon |b|^2 , \quad \varepsilon > 0 . \tag{1.36}$$

Alors, en utilisant la coercivité des A_h (cf. § 1.3.) et la majoration des f_h (cf. (1.26)) on obtient

$$|u_h^{i+1}|_h^2 - |u_h^i|_h^2 + |u_h^{i+1} - u_h^i|_h^2 + (2\alpha - \varepsilon)\, k \|u_k^{i+1}\|_h^2 \leqslant C + 2k\left[j_h(0) - j_h(u_h^{i+1})\right] .$$

Prenons $\varepsilon = \alpha$ et sommons de $i = 0$ à $n \leqslant N$. On obtient [1]

$$|u_h^{n+1}|_h^2 + \sum_{i=0}^{n} |u_h^{i+1} - u_h^i|_h^2 + \alpha \sum_{i=0}^{n} k \|u_h^{i+1}\|_h^2 \leqslant C + |u_{o,h}|_h^2 + \sum_{i=0}^{n} k\left[j_h(0) - j_h(u_h^{i+1})\right]$$

utilisant les majorations (1.6) et (1.11) on a

$$|u_{o,h}|_h^2 \leqslant C \|u_{o,h}\|_h^2 \leqslant C \qquad (h \to 0) .$$

En particulier

$$\alpha \sum_{i=0}^{N-1} k \|u_h^{i+1}\|_h^2 \leqslant C + \sum_{i=0}^{N-1} k\left[j_h(0) - j_h(u_h^{i+1})\right] .$$

D'après le Lemme 1.1 et l'hypothèse faite sur u_h^o on déduit

$$k \|u_h^o\|^2 + \sum_{i=1}^{N} k \|u_h^i\|^2 \leqslant C$$

---

(1) Rappelons encore que d'une ligne sur l'autre la même lettre C peut désigner deux constantes différentes.

ou encore

$$\int_0^T \|u_{h,k}\|_h^2 \, dt \leqslant C \ .$$

De là, il résulte aisément que

$$|q_h u_{h,k}(t)| \leqslant C \ , \text{ constante indépendante de } h,k \text{ et } t$$

En définitive nous avons (1) :

(1.37) $p_h u_{h,k}$ est dans un borné de $L^2(0,T;F)$,

(1.38) $q_h u_{h,k}$ est dans un borné de $L^\infty(0,T;H)$.

1.9. Convergence faible du Schéma I

Nous allons montrer que "$u_{h,k}$" approche la solution (faible) de l'inéquation (1.2). Soit $v \in \mathcal{K}$ et soit $v_{h,k}$ vérifiant les hypothèses (1.12) et (1.13). On remarquera que l'on ne change pas les conclusions (1.12) et (1.13) en imposant $v_{h,k}(0) = u_{o,h}$ donné par (1.10) et (1.11). Portons $v_{h,k}$ dans (1.28) (inéquation de définition du schéma I). En sommant (1.28) on obtient (1.33) qui par le Lemme 1.2 donne (1.34). Comme $v_{h,k}(0) = u_{h,k}(0)$ on obtient

(1.39) $$\int_k^T (\bar{\delta} v_{h,k} + A_h u_{h,k} - f_h \ , \ v_{h,k})_h dt + \int_0^T j_h(v_{h,k}) dt \ \geqslant$$

$$\geqslant \int_k^T (\bar{\delta} v_{h,k} + A_h u_{h,k} - f_h, \ u_{h,k})_h dt + \int_0^T j_h(u_{h,k}) dt \quad ;$$

d'après (1.37) (1.38) il résulte qu'il existe une sous-suite $(h' \ , \ k')$, $h',k' \to 0$, que nous notons encore par (h,k), et $u \in L^\infty(0,T;H)$, $\tilde{u} \in L^2(0,T;F)$ tels que

$$\begin{cases} q_h u_{h,k} \longrightarrow u \ \text{ dans } L^\infty(0,T;H) \text{ faible étoile} \ , \\ p_h u_{h,k} \longrightarrow \tilde{u} \ \text{ dans } L^2(0,T;F) \text{ faible.} \end{cases}$$

D'après l'hypothèse de consistance (1.10) on a alors

$$p_h u_{h,k} \longrightarrow \sigma u \ \text{ dans } L^2(0,T;F) \text{ faible}$$

avec $u \in K$.

On utilise alors toutes les hypothèses de convergence et de consistance pour passer à la limite et l'on obtient

(1.40) $$\int_0^T (v' + Au - f, v) dt + \int_0^T j(v) dt \geqslant \int_0^T (v' + Au - f, u) + \int_0^T j(u) dt$$

---

(1) On dit que $p_h u_{h,k}$ est $L^2(0,T;F)$ stable et que $q_h u_{h,k}$ est $L^\infty(0,T;H)$ stable

de sorte que u est solution (faible) de (1.2) . Les hypothèses requises pour l'unicité étant satisfaites, la convergence reste valable pour toute la suite. On a donc

(1.41)
$$\left|\begin{array}{l} p_h\, u_{h,k} \longrightarrow \sigma u \quad \text{dans } L^2(0,T;F) \text{ faible}, \\ q_h\, u_{h,k} \longrightarrow u \quad \text{dans } L^\infty(0,T;H) \text{ faible étoile.} \end{array}\right.$$

1.10. Convergence forte du schéma I

Nous allons montrer que "$u_{h,k}$" approche la solution (forte) de l'inéquation (1.1) . Pour cela nous supposons d'abord que les conditions d'existence et d'unicité de cette solution sont réalisées. Nous savons alors que les problèmes (1.1) et (1.2) sont équivalents. Soit alors $\tilde{u}_{h,k}$ une suite qui approche u , cette suite étant donnée par les conditions (1.11) et (1.12) . Nous allons évaluer

(1.42)
$$X_{h,k} = \int_0^T (A_h\, u_{h,k} - A_h\, \tilde{u}_{h,k} \,,\, \tilde{u}_{h,k} - u_{h,k})_h dt$$

et montrer que cette quantité tend vers 0 . D'après le Lemme 1.2., en prenant $v_{h,k} = \tilde{u}_{h,k}$ dans l'inéquation (1.34) on obtient

$$\int_k^T (A_h\, u_{h,k} \,,\, u_{h,k})_h dt \leqslant \int_k^T (\bar{\delta}_k\, \tilde{u}_{h,k} \,,\, \tilde{u}_{h,k} - u_{h,k})_h dt +$$
$$+ \int_0^T (A_h\, u_{h,k} \,,\, \tilde{u}_{h,k})_h dt +$$
$$+ \int_0^T \left[j_h(\tilde{u}_{h,k}) - j_h(u_{h,k}) \right] dt -$$
$$- \int_0^T (f_h \,,\, \tilde{u}_{h,k} - u_{h,k})_h dt \,.$$

Mais

$$X_{h,k} = \int_0^T (A_h\, u_{h,k} \,,\, u_{h,k})_h dt - \int_0^T (A_h\, u_{h,k} \,,\, \tilde{u}_{h,k})_h dt - \int_0^T (A_h\, \tilde{u}_{h,k} \,,\, u_{h,k})_h dt +$$
$$+ \int_0^T (A_h\, \tilde{u}_{h,k} \,,\, \tilde{u}_{h,k})_h dt \,.$$

Alors

$$X_{h,k} \leqslant \int_0^T (\bar{\delta}_k\, \tilde{u}_{h,k} \,,\, \tilde{u}_{h,k} - u_{h,k})_h dt + \int_0^T \left[j_h(\tilde{u}_{h,k}) - j_h(u_{h,k}) \right] dt -$$
$$- \int_0^T (f_h \,,\, \tilde{u}_{h,k} - u_{h,k})_h dt - \int_0^T (A_h\, \tilde{u}_{h,k} \,,\, u_{h,k})_h dt + \int_0^T (A_h\, \tilde{u}_{h,k}, \tilde{u}_{h,k})_h d$$

Il résulte alors de

$$\left|\begin{array}{lll} P_h\,\tilde{u}_{h,k} \longrightarrow \sigma u & \text{dans} & L^2(0,T;F) \text{ fort}, \\ q_h\,\bar{\delta}_k\,\tilde{u}_{h,k} \longrightarrow u' & \text{dans} & L^2(0,T;V') \text{ fort}, \\ p_h\,u_{h,k} \longrightarrow \sigma u & \text{dans} & L^2(0,T;F) \text{ faible}, \\ q_h\,u_{h,k} \longrightarrow u & \text{dans} & L^\infty(0,T;H) \text{ faible étoile}, \\ q_h\,\tilde{u}_{h,k} \longrightarrow u & \text{dans} & L^2(0,T;H) \text{ fort}, \end{array}\right.$$

que

$$X_{h,k} \longrightarrow 0$$

et l'on déduit facilement que

$$p_h\,u_{h,k} \longrightarrow \sigma u \quad \text{dans} \quad L^2(0,T;F) \text{ fort}. \qquad (1.43)$$

L'ensemble des résultats obtenus nous permet d'énoncer le

Théorème 1.1.
Les solutions $u_{h,k} = \sum_{i=0}^{N} u_h^i\,\chi_k^i$ du schéma I vérifient

$$\left|\begin{array}{l} p_h\,u_{h,k} \longrightarrow \sigma u \text{ dans } L^2(0,T;F) \text{ faible}, \\ q_h\,u_{h,k} \longrightarrow u \text{ dans } L^\infty(0,T;H) \text{ faible étoile}, \end{array}\right.$$

où u est la solution unique de l'inéquation (1.2). Lorsque cette solution est solution forte unique de (1.1) on a

$$p_h\,u_{h,k} \longrightarrow \sigma u \quad \text{dans} \quad L^2(0,T;F) \text{ fort}. \qquad \blacksquare$$

1.11. Existence et stabilité pour le schéma II

Nous rappelons ce schéma :

$$\left|\begin{array}{l} \text{trouver } u_h^i,\ i=0,1,\ldots,N \text{ tels que} \\ u_h^0 = u_{0,h}, \\ \left(\dfrac{u_h^{i+1}-u_h^i}{k} + A_h\,u_h^{i+\theta} - f_h,\ v_h^{i+1}\right)_h \geq 0 \qquad \forall v_h \in K_h,\ u_h^{i+1} \in K_h \\ \text{avec} \quad u_h^{i+\theta} = u_h^i + \theta\,(u_h^{i+1}-u_h^i),\ \theta \in [0,1]. \end{array}\right. \qquad (1.29)$$

L'existence de u_h^{i+1} est immédiate puisque $(I + \theta\,A_h)$ est coercif, $\forall \theta \in [0,1]$ (pour $\theta = 0$, la coercivité doit être considérée avec la norme :

$|\cdot|_h \geq \frac{1}{S(h)} \|\cdot\|_h$, $S(h) \in]\ 0, +\infty[$, $\forall h$). Nous allons montrer que

$$u_{h,k} = \sum_{i=0}^{N} u_h^i \ \chi_k^i$$

est tel que $p_h\ u_{h,k}$ reste borné dans $L^2(0,T;F)$. Faisant $v=0$ dans (1.29) il vient, en omettant momentanément les indices h ,

$$\frac{1}{k}(u^{i+1} - u^i\ ,\ u^{i+1}) + (A\ u^{i+\theta}\ ,\ u^{i+1}) \leq (f\ ,\ u^{i+1})\ .$$

Utilisant (1.35) on a

$$(1.44)\quad \frac{1}{2k}\ [|u^{i+1}|^2 - |u^i|^2 + |u^{i+1} - u^i|^2] + (A\ u^{i+\theta}\ ,\ u^{i+1}) \leq (f\ ,\ u^{i+1})\ .$$

Or

$$(A\ u^{i+\theta}\ ,\ u^{i+1}) = \theta\ (A\ u^{i+1}\ ,\ u^{i+1}) + (1 - \theta)\ (A\ u^i\ ,\ u^{i+1}) =$$

$$= \theta\ (A\ u^{i+1}\ ,\ u^{i+1}) + (1 - \theta)(A\ u^i\ ,\ u^i) + (1 - \theta)(Au^i\ ,u^{i+1}-$$

et

$$(f\ ,\ u^{i+1}) = (f\ ,\ u^{i+\theta}) + (f\ ,\ u^{i+1} - u^{i+\theta})$$

$$= (1 - \theta)(f\ ,\ u^i) + \theta(f,\ u^{i+1}) + (1 - \theta)(f,\ u^{i+1} - u^i)$$

de sorte que (1.44) donne

$$(1.45)\quad |u^{i+1}|^2 - |u^i|^2 + |u^{i+1} - u^i|^2 + 2k\ [\theta\ (A\ u^{i+1}\ ,\ u^{i+1}) + (1 - \theta)(Au^i\ ,\ u^i)] \leq$$

$$\leq\quad 2k[\theta\ (f\ ,\ u^{i+1}) + (1 - \theta)\ (f,u^i) + (1 - \theta)(f,u^{i+1} - u^i)] -$$

$$- 2k(1 - \theta)(A\ u^i\ ,\ u^{i+1} - u^i)\ .$$

Mais, à partir de (1.36), on a les majorations :

$$|(f\ ,\ u^{i+1})| \leq \|f\|_*\ \|u^{i+1}\| \leq \frac{1}{2\alpha}\ \|f\|_*^2 + \frac{1}{2}(Au^{i+1}\ ,\ u^{i+1})\ ,\ (\varepsilon = \alpha)\ ,$$

$$|(f\ ,\ u^i)| \leq \|f\|_*\ \|u^i\| \leq \frac{1}{\alpha}\ \|f\|_*^2 + \frac{1}{4}\ (Au^i\ ,\ u^i)\quad ,\quad (\varepsilon = \frac{\alpha}{2})\ ,$$

$$|(f\ ,\ u^{i+1} - u^i)| \leq \|f\|_*\ \|u^{i+1} - u^i\| \leq$$

$$\leq \frac{1}{2\varepsilon}\ \|f\|_*^2 + \frac{\varepsilon}{2}\ S(h)^2\ |u^{i+1} - u^i|^2\ ,\ (\varepsilon > 0\ \text{quelconque});$$

$$|(A\,u^i\ ,\ u^{i+1}-u^i)| \leqslant C\,\|u^i\|\ .\ \|u^{i+1}-u^i\| \leqslant \frac{1}{4}(Au^i\ ,\ u^i) + \frac{C'}{\alpha}\,S(h)^2\,|u^{i+1}-u^i|^2\ ,$$

$$(\ \varepsilon = \frac{2}{\alpha})\ ,$$

et reportant dans (1.45) on obtient (nous replaçons les indices h) :

$$(1.46)\quad |u_h^{i+1}|_h^2 - |u_h^i|_h^2 + [1-(\varepsilon+\frac{2}{\alpha}C^2)k\,S(h)^2(1-\theta)]\ |u_h^{i+1}-u_h^i|_h^2 + \\ + (1-\theta)k\,(A_h\,u_h^i\ ,\ u_h^i)_h + \theta k(A_h\,u_h^{i+1},\ u_h^{i+1})_h \leqslant Ck\,\|f_h\|_*^2\ .$$

On introduit alors l'hypothèse suivante, appelée <u>hypothèse de stabilité</u>,

$$(1.47)\qquad 1 - C\,k\,S(h)^2\,(1-\theta) \geqslant \beta > 0\ .$$

Cette hypothèse est évidemment sans objet lorsque $\theta = 1$.

Alors (1.46) donne

$$|u_h^{i+1}|_h^2 - |u_h^i|_h^2 + \beta\,|u_h^{i+1}-u_h^i|_h^2 + \alpha k\,[(1-\theta)\ \|u_h^i\|_h^2 + \theta\,\|u_h^{i+1}\|_h^2] \leqslant Ck\|f_h\|_*^2\ .$$

Avec les hypothèses (1.25)(1.26) sur f_h on a, en sommant de $i=0$ à $n \leqslant N$,

$$(1.48)\quad |u_h^{n+1}|_h^2 + \sum_{i=0}^{n}\beta\,|u_h^{i+1}-u_h^i|_h^2 + \alpha\sum_{i=0}^{n} k\,[(1-\theta)\,\|u_h^i\|_h^2 + \theta\,\|u_h^{i+1}\|_h^2] \leqslant C.$$

D'où l'on déduit que

$$(1.49)\quad p_h\,u_{h,k} \text{ est dans un borné de } L^2(0,T;F)\ ,$$

$$(1.50)\quad q_h\,u_{h,k} \text{ est dans un borné de } L^\infty(0,T;H)\ ,$$

$$(1.51)\quad \sum_{i=0}^{N}\beta\,|u_h^{i+1}-u_h^i|_h^2 \leqslant C\ .$$

1.12. <u>Convergence faible du schéma II.</u>

On prend $v \in \mathcal{K}$ et $v_{h,k}$ comme au § 1.9. On porte $v_{h,k}$ dans (1.29). Compte tenu de la Remarque 1.2 (§ 1.7.) on obtient (avec $j=0$)

$$(1.52)\quad \int_k^T (\delta_k v_{h,k} + A_h\,u_{h,k}^\theta - f_h\ ,\ v_{h,k}-u_{h,k})_h\,dt \geqslant \int_k^T (\delta\ v_{h,k}\ ,\ u_{h,k})dt + \int_k^T (A_h\,u_{h,k}^\theta\ ,u_{h,k})dt \\ + \int_k^T (f_h,\ u_{h,k})dt$$

où

$$u_{h,k}^\theta(t) = u_{h,k}(t-k) + \theta\,[u_{h,k}(t) - u_{h,k}(t-k)]\ .$$

Mais d'après (1.49)(1.50) et les hypothèses de consistance (1.10) on peut extraire une sous-suite encore notée $u_{h,k}$ telle que :

$q_h\, u_{h,k} \longrightarrow u$ dans $L^\infty(0,T;H)$ faible étoile,

$p_h\, u_{h,k} \longrightarrow u$ dans $L^2(0,T;F)$ faible,

avec $u \in K$.

En passant à la limite dans (1.52) on obtient

$$(1.53)\quad \int_0^T (v' + Au - f\ ,\ v - u)dt \geqslant \liminf \int_k^T (A_h\, u^\theta_{h,k}\ ,\ u_{h,k} - u^\theta_{h,k})_h dt.$$

Il reste à montrer que le second membre est $\geqslant 0$.

On a la majoration

$$(1.54)\quad Y_{h,k} = \left| \int_k^T (A_h\, u^\theta_{h,k}\ ,\ u_{h,k} - u^\theta_{h,k})_h dt \right| = \left| \sum_{i=0}^N k(A_h\, u_h^{i+\theta}\ ,\ u_h^{i+1} - u_h^{i+\theta})_h \right| =$$

$$= \sum_{i=0}^N k(1-\theta)\ \left|(A_h\, u_h^{i+\theta}\ ,\ u_h^{i+1} - u_h^i)_h\right| \leqslant$$

$$\leqslant (1-\theta)\ S(h) \sum_{i=0}^N \sqrt{k}\ \|u_h^{i+\theta}\|_h \cdot \sqrt{k}\,|u_h^{i+1} - u_h^i| \leqslant^{(1)}$$

$$\leqslant (1-\theta)\ S(h)\ \Big(\sum_{i=0}^N k \|u_h^{i+\theta}\|_h^2\Big)^{\frac{1}{2}} \Big(\sum_{i=0}^N |u_h^{i+1} - u_h^i|_h^2\Big)^{\frac{1}{2}} \sqrt{k}\ .$$

En utilisant (1.49) (1.51) et l'hypothèse de stabilité (1.47) il vient

$$Y_{h,k} \longrightarrow 0\ ,\quad h,k \longrightarrow 0$$

d'où

$$\int_0^T (v' + Au - f,\ v-u)dt \geqslant 0$$

et u est bien solution "faible" de(1.2.). ∎

1.13. Convergence forte du schéma II

On raisonne comme au § 1.10 avec les mêmes hypothèses. Nous allons montrer que $X_{h,k}$ donné par (1.42) tend vers 0. D'après le Lemme 1.2 , en prenant $v_{h,k} = \tilde{u}_{h,k}$ dans l'inéquation (1.34) (avec $j=0$ et compte tenu de la Remarque 1.2) , (1.34) étant écrit avec $A_h\, u_{h,k}(t)$ remplacé par $A_h\, u^\theta_{h,k}(t)$, on obtient :

---

(1) $\left(\sum a_i\, b_i\right)^2 \leqslant \left(\sum a_i^2\right)\left(\sum b_i^2\right)$.

$$\int_k^T (\delta_k \tilde{u}_{h,k}(t) + A_h u_{h,k}(t+1) - (1-\theta) A_h(u_{h,k}(t) - u_{h,k}(t-1)) - f_h \, , \, \tilde{u}_{h,k} - u_{h,k})_h$$

$$\geqslant 0$$

d'où

$$\int_k^T (A_h(u_{h,k} - \tilde{u}_{h,k}), u_{h,k} - \tilde{u}_{h,k})_h dt \leqslant \int_k^T (\delta_k \tilde{u}_{h,k} - f_h \, , \, \tilde{u}_{h,k} - u_{h,k})_h dt +$$

$$+ |\theta - 1| \int_k^T (A_h(u_{h,k}(t) - u_{h,k}(t-k)), \tilde{u}_{h,k} - u_{h,k})_h dt$$
$$+ O(k,k).$$

En vertu des résultats du § 1.10 il reste à prouver que

$$(1.55) \qquad (1-\theta) \left| \sum_{i=0}^{N} k(A_h(u_h^{i+1} - u_h^i) \, , \, \tilde{u}_h^{i+1} - u_h^{i+1}) \right| \to 0.$$

En majorant comme en (1.54), en utilisant (1.51), la convergence de $\tilde{u}_{h,k}$ (donc cette suite est bornée), et (1.49) on déduit (1.55) et la convergence est établie.
Nous pouvons énoncer le

Théorème 1.2.

Si $\theta \in [0,1]$ et si $h,k \to 0$ avec

$$(1.47) \qquad \left| \quad (1-\theta)\, k\, S(h)^2 \to 0 \right.$$

alors les solutions $u_{h,k} = \sum u_h^i \chi_k^i$ des schémas II vérifient :

$$\left| \quad p_h u_{h,k} \longrightarrow \sigma u \text{ dans } L^2(0,T;F) \text{ faible}, \right.$$

$$\left| \quad q_h u_{h,k} \longrightarrow u \text{ dans } L^\infty(0,T;H) \text{ faible étoile}, \right.$$

où u est la solution unique de l'inéquation (1.2). Lorsque cette solution est solution forte unique de (1.1) on a :

$$\left| \quad p_h u_{h,k} \longrightarrow \sigma u \text{ dans } L^2(0,T;F) \text{ fort}. \right.$$

(L'hypothèse (1.47) est sans objet si $\theta = 1$). ■

Remarque 1.3.

Lorsque h et k peuvent tendre vers 0 indépendamment l'un de l'autre on dit que le schéma est inconditionnellement stable. C'est donc le cas ici avec $\theta = 1$. De même, le schéma I étudié précédemment est inconditionnellement stable. ■

1.11. Existence et stabilité pour le schéma III (schéma découplé)

Rappelons le schéma :

trouver $u_h^{i+\frac{1}{2}}$ et u_h^i , $i=0,1,\dots,N$ tels que

$$u_h^0 = u_{0,h} \,, \tag{1.56}$$

$$\frac{u_h^{i+\frac{1}{2}} - u_h^i}{k} + A_h(\theta^- u_h^i + \theta^* u_h^{i+\frac{1}{2}} + \theta^+ u_h^{i+1}) - f_h = 0 \,, \tag{1.57}$$

$$u_h^{i+1} = P_{K_h}\left[u_h^{i+\frac{1}{2}}\right] \,, \tag{1.58}$$

$$\theta^- , \theta^* , \theta^+ \geqslant 0 \,, \quad \theta^- + \theta^* + \theta^+ = 1 \,.$$

Par définition de P_{K_h}, (1.58) est équivalent à

$$(u_h^{i+1} - u_h^{i+\frac{1}{2}} \,, v_h - u_h^{i+1})_h \geqslant 0 \,, \quad \forall v_h \in K_h \,, \; u_h^{i+1} \in K_h \,. \tag{1.59}$$

Nous allons établir l'existence de u_h^{i+1} .

Pour la commodité nous omettrons l'indice h. L'équation (1,57) de définition de $u_h^{i+\frac{1}{2}}$ s'écrit

$$u^{i+\frac{1}{2}} - u^i + k\left[A\left(\theta^- u^i + \theta^* u^{i+\frac{1}{2}} + \theta^+ u^{i+1}\right) - f\right] = 0 \,,$$

soit

$$(I + k\,\theta^* A)\, u^{i+\frac{1}{2}} = (I - k\,\theta^- A)u^i - k\,\theta^+ Au^{i+1} + kf \,.$$

Posons

$$\mathcal{A} = (I + k\,\theta^* A) \,.$$

Puisque la matrice A est inversible alors $\mathcal{A}^{-1}$ existe pour $\theta^* \geqslant 0$, et

$$u^{i+\frac{1}{2}} = \mathcal{A}^{-1}(I - k\,\theta^- A)\, u^i - k\,\theta^+ \mathcal{A}^{-1} Au^{i+1} + k\,\mathcal{A}^{-1} f \,.$$

On peut écrire (1.59) en remplaçant $u^{i+\frac{1}{2}}$ par sa valeur, soit

$$(I + k\,\theta^+ \mathcal{A}^{-1} A)\, u^{i+1} - \mathcal{A}^{-1}(I - k\,\theta^- A)u^i - k\,\mathcal{A}^{-1} f \,,\, v - u^{i+1}) \geqslant 0, \; \forall v \in K$$

L'existence découle alors de la coercivité de l'opérateur qui prémultiplie u^{i+1} (cf. Ch.I).

Cherchons donc s'il existe une constante $\mu > 0$ telle que

$$((I + k\ \theta^+ a^{-1} A)v,v) \geqslant \mu |v|^2 .$$

Pour cela il suffit que

$$(a^{-1} A v,v) \geqslant 0, \qquad \forall v \in V_h .$$

Or

$$\begin{aligned}(I + k\ \theta^* A)^{-1} A &= (I + k\ \theta^* A)^{-1} (A^{-1})^{-1} = \\ &= [A^{-1}(I + k\ \theta^* A)]^{-1} = \\ &= [A^{-1} + k\ \theta^* I]^{-1} = \\ &= [(I + k\ \theta^* A) A^{-1}]^{-1} = \\ &= (A^{-1})^{-1}(I + k\ \theta^* A)^{-1} = \\ &= A(I + k\ \theta^* A)^{-1} .\end{aligned}$$

Il faut donc regarder si

$$(A\ a^{-1} v,v) \geqslant 0 , \qquad \forall v \in V_h .$$

Posons $w = a^{-1} v$. Alors

$$\begin{aligned}(A\ a^{-1} v,v) &= (A w , a w) = \\ &= (A w , (I + k\ \theta^* A)w) = \\ &= (A w,w) + k\ \theta^* (Aw , Aw) \geqslant \\ &\geqslant \alpha \|w\|^2 + k\ \theta^* \|Aw\|^2 \\ &\geqslant \alpha \|w\|^2 .\end{aligned}$$

L'existence est donc établie. ∎

Nous allons maintenant étudier la stabilité du schéma III. Ici encore nous omettrons l'indice h.

Posons

$$u^{i+\theta} = \theta^- u_h^i + \theta^* u_h^{i+\frac{1}{2}} + \theta^+ u_h^{i+1} . \tag{1.60}$$

Multiplions (1.57) par $u^{i+\frac{1}{2}}$. Il vient

$$\frac{1}{k}(u^{i+\frac{1}{2}} - u^i , u^{i+\frac{1}{2}}) + (Au^{i+\theta} , u^{i+\frac{1}{2}}) = (f, u^{i+\frac{1}{2}}) . \tag{1.61}$$

On obtient (cf. (1.35)):

(1.62) $\frac{1}{2k}\left[|u^{i+\frac{1}{2}}|^2 - |u^i|^2 + |u^{i+\frac{1}{2}} - u^i|^2\right] + \theta^-(Au^i , u^{i+\frac{1}{2}}) + \theta^*(Au^{i+\frac{1}{2}} , u^{i+\frac{1}{2}}) +$

$$\theta^+(Au^{i+1} , u^{i+\frac{1}{2}}) = (f , u^{i+\frac{1}{2}})$$

d'où

(1.63) $\frac{1}{2k}\left[|u^{i+\frac{1}{2}}|^2 - |u^i|^2 + |u^{i+\frac{1}{2}} - u^i|^2\right] + \theta^-(Au^i , u^i) + \theta^*(Au^{i+\frac{1}{2}} , u^{i+\frac{1}{2}}) + \theta^+(Au^{i+1} , u^{i+1}) =$

$$= (f,u^{i+\frac{1}{2}}) - \theta^-(Au^i , u^{i+\frac{1}{2}} - u^i) - \theta^+(Au^{i+1} , u^{i+\frac{1}{2}} - u^{i+1}) =$$

$$= (f,u^{i+\frac{1}{2}} - \theta^- u^i - \theta^* u^{i+\frac{1}{2}} - \theta^+ u^{i+1}) + (f, u^{i+\theta}) -$$

$$- \theta^-(Au^i , u^{i+\frac{1}{2}} - u^i) - \theta^+(Au^{i+1} , u^{i+\frac{1}{2}} - u^{i+1}) =$$

$$= \theta^-(f , u^{i+\frac{1}{2}} - u^i) + \theta^+(f, u^{i+\frac{1}{2}} - u^{i+1}) + (f, u^{i+\theta})$$

$$- \theta^-(Au^i , u^{i+\frac{1}{2}} - u^i) - \theta^+(Au^{i+1} , u^{i+\frac{1}{2}} - u^{i+1}) .$$

Or, en utilisant (1.36), il vient

(1.64) $$|(f,u^i)| \leq \frac{1}{2\varepsilon_1} \|f\|_*^2 + \frac{\varepsilon_1}{2\alpha}(Au^i , u^i) \quad , \quad (\varepsilon = \alpha) ,$$

(1.65) $$|(f,u^{i+\frac{1}{2}})| \leq \frac{1}{2\varepsilon_2} \|f\|_*^2 + \frac{\varepsilon_2}{2\alpha}(Au^{i+\frac{1}{2}} , u^{i+\frac{1}{2}}) ,$$

(1.66) $$|(f,u^{i+1})| \leq \frac{1}{2\varepsilon_3} \|f\|_*^2 + \frac{\varepsilon_3}{2\alpha}(Au^{i+1} , u^{i+1}) ,$$

(1.67) $$|(f,u^{i+\frac{1}{2}} - u^i)| \leq \frac{1}{2\varepsilon_4} \|f\|_*^2 + \frac{\varepsilon_4}{2} S(h)^2 |u^{i+\frac{1}{2}} - u^i|^2 ,$$

(1.68) $$|(f,u^{i+\frac{1}{2}} - u^{i+1})| \leq \frac{1}{2\varepsilon_5} \|f\|_*^2 + \frac{\varepsilon_5}{2} S(h)^2 |u_h^{i+\frac{1}{2}} - u^{i+1}|^2 ,$$

(1.69) $$|(Au^i,u^{i+\frac{1}{2}} - u^i)| \leq \frac{1}{2\varepsilon_6\alpha}(Au^i,u^i) + \varepsilon_6 C S(h)^2 |u^{i+\frac{1}{2}} - u^i|^2 ,$$

(1.70) $$|(Au^{i+1},u^{i+\frac{1}{2}} - u^{i+1})| \leq \frac{1}{2\varepsilon_7\alpha}(Au^{i+1},u^{i+1}) + \varepsilon_7 C S(h)^2 |u^{i+\frac{1}{2}} - u^{i+1}|^2 .$$

De même en portant la valeur de $u^{i+\frac{1}{2}}$ donnée par (1.57) dans (1.59) et en faisant $v = 0$ dans (1.59) on obtient

(1.71) $\frac{1}{2k}\left[|u^{i+1}|^2 - |u^i|^2 + |u^{i+1} - u^i|^2\right] + \theta^-(Au^i,u^{i+1}) + \theta^*(Au^{i+\frac{1}{2}} , u^{i+1}) +$

$$+ \theta^+(Au^{i+1} , u^{i+1}) \leq (f, u^{i+1})$$

soit

$$(1.72)\quad \frac{1}{2k}\left[|u^{i+1}|^2 - |u^i|^2 + |u^{i+1}-u^i|^2\right] + \theta^-(Au^i,u^i) + \theta^*(Au^{i+\frac12}, u^{i+\frac12}) + \theta^+(Au^{i+1}, u^{i+1}) \leqslant$$

$$\leqslant \theta^-(f,u^{i+1}-u^i) + \theta^*(f,u^{i+1}-u^{i+\frac12}) + (f,u^{i+\theta}) - \theta^-(Au^i, u^{i+1}-u^i) - \theta^*(Au^{i+\frac12}, u^{i+1}-u^{i+\frac12}).$$

Mais en utilisant (1.36) il vient

$$(1.73)\quad |(f,u^{i+1}-u^i)| \leqslant \frac{1}{2\varepsilon_8}\|f\|_*^2 + \frac{\varepsilon_8}{2} S(h)^2 |u^{i+1}-u^i|^2 ,$$

$$(1.74)\quad |(f,u^{i+1}-u^{i+\frac12})| \leqslant \frac{1}{2\varepsilon_9}\|f\|_*^2 + \frac{\varepsilon_9}{2} S(h)^2 |u^{i+1}-u^{i+\frac12}|^2 ,$$

$$(1.75)\quad |(Au^i, u^{i+1}-u^i)| \leqslant \frac{1}{2\varepsilon_{10}\alpha}(Au^i,u^i) + \varepsilon_{10} C S(h)^2 |u^{i+1}-u^i|^2 ,$$

$$(1.76)\quad |(Au^{i+\frac12}, u^{i+1}-u^{i+\frac12})| \leqslant \frac{1}{2\varepsilon_{11}\alpha}(Au^{i+\frac12}, u^{i+\frac12}) + \varepsilon_{11} C S(h)^2 |u^{i+1}-u^{i+\frac12}|^2 ,$$

En utilisant les majorations (1.64) (1.65) (1.66) (1.67) (1.68) reportées dans (1.63) on obtient, après multiplication par k ,

$$(1.77)\quad |u^{i+\frac12}|^2 - |u^i|^2 + [1 - \theta^-(\varepsilon_4 + C\varepsilon_6) k S(h)^2] \cdot |u^{i+\frac12}-u^i|^2 + \theta^-(\alpha - \varepsilon_1 - \frac{1}{\varepsilon_6}) k \|u^i\|^2 + \theta^*(\alpha - \varepsilon_2) k \|u^{i+\frac12}\|^2 + \theta^+(\alpha - \varepsilon_3 - \frac{1}{\varepsilon_7}) k \|u^{i+1}\|^2 \leqslant (\frac{\theta^-}{\varepsilon_1} + \frac{\theta^*}{\varepsilon_2} + \frac{\theta^+}{\varepsilon_3} + \frac{\theta^-}{\varepsilon_4} + \frac{\theta^+}{\varepsilon_5}) k \|f\|_*^2 + \theta^+ [\varepsilon_5 + C\varepsilon_7] k S(h)^2 |u^{i+\frac12} - u^{i+1}|^2 .$$

De même en utilisant les majorations (1.64) (1.65) (1.66) (1.73) (1.74) (1.75) (1.76) reportées dans (1.72) on obtient , après multiplication par k ,

$$(1.78)\quad |u^{i+1}|^2 - |u^i|^2 + [1 - \theta^-(\varepsilon_8 + C\varepsilon_{10}) k S(h)^2] \cdot |u^{i+1}-u^i|^2 + \theta^-(\alpha - \varepsilon_1 - \frac{1}{\varepsilon_{10}}) k \|u^i\|^2 + \theta^*(\alpha - \varepsilon_2 - \frac{1}{\varepsilon_{11}}) k \|u^{i+\frac12}\|^2 + \theta^+(\alpha - \varepsilon_3) k \|u^{i+1}\|^2 \leqslant (\frac{\theta^-}{\varepsilon_1} + \frac{\theta^*}{\varepsilon_2} + \frac{\theta^+}{\varepsilon_3} + \frac{\theta^-}{\varepsilon_8} + \frac{\theta^*}{\varepsilon_9}) k \|f\|_*^2 + \theta^*[\varepsilon_9 + C\varepsilon_{11}] k S(h)^2 |u^{i+\frac12} - u^{i+1}|^2 .$$

Par ailleurs, comme $0 \in K$, on a

$$(u^{i+1} - u^{i+\frac{1}{2}} , 0 - u^{i+1}) \geqslant 0$$

d'où

$$|u^{i+1}|^2 - |u^{i+\frac{1}{2}}|^2 + |u^{i+1} - u^{i+\frac{1}{2}}|^2 \leqslant 0$$

et

$$|u^{i+1}|^2 + |u^{i+1} - u^{i+\frac{1}{2}}|^2 \leqslant |u^{i+\frac{1}{2}}|^2 \quad . \tag{1.79}$$

En tenant compte de cette minoration et en ajoutant (1.77) et (1.78) on obtient

$$\begin{aligned} 2|u^{i+1}|^2 - 2|u^i|^2 + |u^{i+1} - u^{i+\frac{1}{2}}|^2 & [1 - (\theta^+(\varepsilon_5 + C\varepsilon_7) + \theta^*(\varepsilon_9 + C\varepsilon_{11}))k\, S(h)^2] + \\ + [1 - \theta^-(\varepsilon_4 + C\varepsilon_6)\, k\, S(h)^2] & \quad |u^{i+\frac{1}{2}} - u^i|^2 + \\ + [1 - \theta^-(\varepsilon_8 + C\varepsilon_{10}) k\, S(h)^2] & \quad |u^{i+1} - u^i|^2 + \\ + \theta^-(2\alpha - 2\varepsilon_1 - \frac{1}{\varepsilon_6} - \frac{1}{\varepsilon_{10}})k \|u^i\|^2 & + \\ + \theta^*(2\alpha - 2\varepsilon_2 - \frac{1}{\varepsilon_{11}})k \|u^{i+\frac{1}{2}}\|^2 & + \\ + \theta^+(2\alpha - 2\varepsilon_3 - \frac{1}{\varepsilon_7})k \|u^{i+1}\|^2 & \leqslant \\ \leqslant \quad (\frac{2\theta^-}{\varepsilon_1} + \frac{2\theta^*}{\varepsilon_2} + \frac{2\theta^+}{\varepsilon_3} + \frac{\theta^-}{\varepsilon_4} + \frac{\theta^-}{\varepsilon_8} + \frac{\theta^*}{\varepsilon_9} + \frac{\theta^+}{\varepsilon_5})k \|f\|_*^2 & \quad . \end{aligned} \tag{1.80}$$

En prenant ε_i , $i = 1, \ldots 11$, constantes > 0 convenables, en introduisant la condition de stabilité

$$1 - C\, k\, S(h)^2 \geqslant \beta > 0 \tag{1.81}$$

et en posant

$$u_{h,k}^- = \sum_{i=0}^{N} u_h^i \, \chi_h^i \, , \tag{1.82}$$

$$u_{h,k}^* = \sum_{i=0}^{N} u^{i+\frac{1}{2}} \, \chi_k^i \, , \tag{1.83}$$

$$u_{h,k}^+ = \sum_{i=0}^{N} u^{i+1} \, \chi_k^i \, , \tag{1.84}$$

on obtient, par sommation de (1.80),

(1.85) $\theta^{-} p_h u^{-}_{h,k}$, $\theta^{*} p_h u^{*}_{h,k}$, $\theta^{+} p_h u^{+}_{h,k}$ sont bornés dans $L^2(0,T;F)$,

(1.86) $q_h u^{-}_{h,k}$, $q_h u^{+}_{h,k}$ sont bornés dans $L^{\infty}(0,T;H)$,

(1.87)
$$\sum_{i=0}^{N} \beta \left| u^{i+1} - u^{i+\frac{1}{2}} \right|^2 \leqslant C ,$$
$$\sum_{i=0}^{N} \beta \left| u^{i+\frac{1}{2}} - u^{i} \right|^2 \leqslant C ,$$
$$\sum_{i=0}^{N} \beta \left| u^{i+1} - u^{i} \right|^2 \leqslant C .$$

Il suffit de reporter ces résultats dans (1.77) pour obtenir

(1.88) $q_h u^{*}_{h,k}$ est borné dans $L^{\infty}(0,T;H)$. ∎

A première vue, il semblerait que ces schémas soient toujours conditionnellement stables et convergents. Nous allons voir cependant que cela n'est pas vrai et qu'il existe toute une sous-famille de ces schémas qui est inconditionnellement stable (lorsque $K_h = V_h$). Pour cela nous reprenons (1.57) que nous multiplions par $u^{i+\theta}$ donné par (1.60). Il vient

(1.89)
$$\frac{1}{k}(u^{i+\frac{1}{2}} - u^{i} , u^{i+\theta}) + (Au^{i+\theta} , u^{i+\theta}) = (f, u^{i+\theta}) .$$

En utilisant (1.35) le premier terme donne

$$(u^{i+\frac{1}{2}} - u^{i} , u^{i}) = \left|u^{i+\frac{1}{2}}\right|^2 - \left|u^{i}\right|^2 - \left|u^{i+\frac{1}{2}} - u^{i}\right|^2 ,$$
$$(u^{i+\frac{1}{2}} - u^{i} , u^{i+\frac{1}{2}}) = \left|u^{i+\frac{1}{2}}\right|^2 - \left|u^{i}\right|^2 + \left|u^{i+\frac{1}{2}} - u^{i}\right|^2 ,$$
$$(u^{i+\frac{1}{2}} - u^{i} , u^{i+1}) = (u^{i+\frac{1}{2}} - u^{i+1} , u^{i+1}) + (u^{i+1} - u^{i} , u^{i+1}) .$$

Puisque $0 \in K$, on a

$$(u^{i+\frac{1}{2}} - u^{i+1} , u^{i+1}) \geqslant 0$$

donc

$$(u^{i+\frac{1}{2}} - u^{i} , u^{i+1}) \geqslant \left[\left|u^{i+1}\right|^2 - \left|u^{i}\right|^2 + \left|u^{i+1} - u^{i}\right|^2 \right] .$$

On obtient alors

$$(1.90)\quad |u^{i+\frac{1}{2}}|^2 - |u^i|^2 + (\theta^* - \theta^-)\,|u^{i+\frac{1}{2}} - u^i|^2 + \theta^+|u^{i+1} - u^i|^2 + k(2\alpha - \varepsilon)\|u^{i+\theta}\|^2 \leq \frac{k}{\varepsilon}\|f\|_*^2 .$$

On peut dès maintenant reconnaître que le cas $\theta^* > \theta^-$ semble conduire à une stabilité inconditionnelle. Nous allons examiner cela de plus près.

En utilisant les majorations

$$(1.91)\quad |u^{i+1}|^2 + |u^{i+1} - u^{i+\frac{1}{2}}|^2 \leq |u^{i+\frac{1}{2}}|^2 ,$$

$$(1.92)\quad |u^{i+1} - u^{i+\frac{1}{2}}| \leq |u^{i+\frac{1}{2}} - u^i| ,$$

$$(1.93)\quad |u^{i+\frac{1}{2}} - u^i| \leq |u^{i+1} - u^i| + |u^{i+1} - u^{i+\frac{1}{2}}| ,$$

$$(1.94)\quad |u^{i+1} - u^i| \leq |u^{i+\frac{1}{2}} - u^i| ,$$

on déduit, avec $\varepsilon>0$, fixe, aussi petit que l'on veut, $\varepsilon \leq \alpha$, si $\theta^+ + \theta^* \geq \theta^- + \varepsilon$ alors "inconditionnellement ",

$$(1.95)\quad p_h\, u_{h,k}^{\theta} \text{ est borné dans } L^2(0,T;F)$$

$$(1.96)\quad \text{où l'on a posé } u_{h,k}^{\theta} = \sum_{i=0}^{N} u_h^{i+\theta}\, \chi_h^i) ,$$

$$(1.97)\quad q_h\, u_{h,k}^{*} \text{ est borné dans } L^\infty(0,T;H) ,$$

$$(1.98)\quad q_h\, u_{h,k}^{+} \text{ est borné dans } L^\infty(0,T;H) ,$$

$$(1.99)\quad \sum_{i=0}^{N} |u^{i+1} - u^i|^2 \leq C ,$$

$$(1.100)\quad \sum_{i=0}^{N} |u^{i+\frac{1}{2}}\, u^i|^2 \leq C ,$$

$$(1.101)\quad \sum_{i=0}^{N} |u^{i+1} - u^{i+\frac{1}{2}}|^2 \leq C ,$$

mais il n'y a aucune raison (sauf si $K_h = V_h$) pour que $p_h\, u_{h,k}^{+}$ soit dans un ensemble borné de $L^2(0,T;F)$. ■ Nous allons regarder ce que donne (1.90) lorsque $\theta^* < \theta^-$. Pour cela nous devons majorer $|u^{i+\frac{1}{2}} - u^i|$. On a

$$u^{i+\frac{1}{2}} - u^i = -k(Au^{i+\theta} - f)$$

d'où

$$|u^{i+\frac{1}{2}} - u^i|^2 = k(f, u^{i+\frac{1}{2}} - u^i) - k(Au^{i+\theta} , u^{i+\frac{1}{2}} - u^i)$$

et

$$|u^{i+\frac{1}{2}} - u^i| \leqslant k \|f\|_* + C\, k\, S(h) \| u^{i+\theta} \|$$

soit

$$(1.102) \qquad |u^{i+\frac{1}{2}} - u^i|^2 \leqslant (1+\tfrac{1}{\varepsilon'})\, k^2 \|f\|_*^2 + (1+\varepsilon')\, C\, k^2\, S(h)^2 \| u^{i+\theta} \|^2 , \ \varepsilon' > 0.$$

En reportant (1.102) dans (1.90) on obtient

$$(1.103) \qquad |u^{i+\frac{1}{2}}|^2 - |u^i|^2 + \theta^+ |u^{i+1} - u^i|^2 +$$

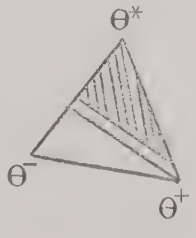
$$+ k\,[2\alpha - \varepsilon - (\theta^- - \theta^*)(1+\varepsilon')\, C\, k\, S(h)^2] \| u^{i+\theta} \|^2 \leqslant$$

$$\leqslant k\, [\tfrac{1}{\varepsilon} + k(\theta^- - \theta^*)(1+\tfrac{1}{\varepsilon'})] \| f \|_*^2 .$$

Avec β comme en (1.81) on a donc

$$|u^{i+\frac{1}{2}}|^2 - |u^i|^2 + \theta^+ |u^{i+1} - u^i|^2 + \beta k \| u^{i+\theta} \|^2 \leqslant Ck \| f \|_*^2 .$$

Ainsi, si $\theta^- > \theta^*$, il résulte de (1.90) que le schéma est conditionnellement stable.



1.12. Convergence faible du schéma III

Nous allons d'abord montrer que les fonctions $u_{h,k}^-$, $u_{h,k}^*$, $u_{h,k}^+$, $u_{h,k}^\theta$ (définies par (1.82)(1.83)(1.84)(1.96)) tendent vers la même limite. Soit $v \in V$ [1] et soit $\varphi(t) \in \mathcal{C}^0(0,T)$ espace des fonctions scalaires continues sur $[0,T]$. Soit $v_h = V_h$ tel que

$$(1.104) \qquad q_h\, v_h \longrightarrow v \text{ dans } H \text{ fort} ,$$

$$(1.105) \qquad p_h\, v_h \longrightarrow \sigma v \text{ dans } F \text{ fort} ,$$

(ce qui est licite si l'on a les hypothèses de convergence (1.11) avec K_h remplacé par V_h). L'équation (1.57) s'écrit alors, après multiplication par v_h ,

$$(1.106) \qquad (q_h\, u_{h,k}^* - q_h\, u_{h,k}^- , q_h\, v_h) + k\, a_{h,k}(t; u_{h,k}^\theta , v_h) = k(f_{h,k} , q_h\, v_h) .$$

Multiplions par $\varphi(t)$ et intégrons (1.106), il vient

$$(1.107) \qquad \int_0^{T-k} (q_h\, u_{h,k}^* - q_h\, u_{h,k}^- , \varphi\, q_h\, v_h)dt + k \int_0^{T-k} a_{h,k}(t; u_{h,k}^\theta , \varphi(t)\, v_h)dt =$$

$$= k \int_0^{T-k} (f_{h,k} , \varphi v_h)dt .$$

(1) On peut se restreindre à ne prendre v que dans un espace γ dense dans V .

D'après les hypothèses (1.15) et (1.26) on obtient

$$(1.108)\quad \left| \int_0^{T-k} (q_h\, u_{h,k}^{*} - q_h\, u_{h,k}^{-}\ ,\ \varphi(t)\, q_h\, v_h)dt \right| \leqslant$$

$$\leqslant k\, |\varphi|_{L^2(0,T)} \{ M \| p_h\, u_{h,k}^{\theta} \|_{L^2(0,T;F)} \| p_h\, v_h \|_F + C \| p_h\, v_h \|_F \} .$$

Or les termes de l'accolade restent bornés lorsque h et $k \to 0$. Donc le premier membre tend vers 0 lorsque h et $k \to 0$. D'après la linéarité de q_h on a

$$q_h\ \varphi v_h = \varphi q_h\, v_h \ ,$$

or d'après (1.104) on a

$$q_h \varphi . v_h \ \to\ \varphi\, v \quad \text{dans } H \text{ fort}, \ \forall t \in [0,T] \ ,$$

de plus

$$|q_h\ \varphi v_h| \ \leqslant\ C\, |\varphi| \qquad (C = |v| + C)$$

et $|\varphi|$ étant intégrale, le Théorème de Lebesgue donne (1)

$$q_h\ \varphi\, v_h \to \varphi v \quad \text{dans } L^p(0,T;H) \text{ fort}, \ 1 \leqslant p \leqslant \infty \ ,$$

d'où, ici, la convergence dans $L^1(0,T;H)$ fort. Comme $u_{h,k}^{*}$ et $u_{h,k}^{-}$ sont dans un borné de $L^{\infty}(0,T;H)$ on a, en extrayant une sous-suite que nous indicerons encore par h et k ,

$$\int_0^{T-k} (q_h\, u_{h,k}^{*} - q_h\, u_{h,k}^{-}\ ,\ \varphi(t)\, q_h\, v_h)dt \to \int_0^{T} (u^{*} - u^{-}\ ,\ \varphi(t)\, v)dt$$

où u^{*} (resp. u^{-}) est limite dans $L^{\infty}(0,T;H)$ faible étoile de

$$q_h\, u_{h,k}^{*} \ (\text{resp. } q_h\, u_{h,k}^{-}) \ .$$

De la convergence vers 0 du premier terme de (1.108) on déduit que :

$$(1.109)\qquad \int_0^{T} (u^{*} - u^{-}\ ,\ \varphi v)dt = 0 \quad , \qquad \forall\, \varphi \in \mathcal{C}(0,T) \ , \ \forall v \in V \ .$$

Comme l'espace des fonctions sommes finies de $\varphi_i \otimes v_i$, avec $\varphi_i \in \mathcal{C}(0,T)$, $v_i \in H$ est dense dans $L^2(0,T;H)$, et que V est dense dans H on a :

$$(1.110)\qquad u^{*} = u^{-} \quad \text{dans } L^2(0,T;H) \ ,$$

---

(1) cf. par exemple TEMAM [2] Lemme 16.4, p. 247 .

(1.111) $$u^* = u^- \quad \text{p.p. dans } [0,T] \ .$$ (1)

Etudions maintenant la convergence de $u^+_{h,k}$ (cf.(1.84). On a

$$u^+_{h,k}(t) = u^-_{h,k}(t+k) \ .$$

En conséquence, pour une sous-suite de la sous-suite considérée :

$$u^+_{h,k} \to u^+ = u^- \text{ dans } L^2(0,T;H) \text{ et p.p. dans } [0,T] \ .$$

Par ailleurs comme $u^+_{h,k} \in K_h$ on a, avec l'hypothèse de consistance,

$$u^+ \in K \ .$$

Comme d'autre part

$$u^\theta_{h,k} = \theta^- u^-_{h,k} + \theta^* u^*_{h,k} + \theta^+ u^+_{h,k}$$

on a

(1.112) $$u^- = u^* = u^+ = u^\theta = u \ . \quad \blacksquare$$

Nous allons montrer que u est solution faible de (1.2). Pour cela on commence comme en (1.52)(mais $u^\theta_{h,k}$ étant donné par (1.62) et (1.96)). En utilisant le fait que

$$p_h \, u^\theta_{h,k} \text{ est borné dans } L^2(0,T;F)$$

on obtient, par extraction d'une sous-suite,

$$\int_0^T (v' + Au - f, \, v-u)dt \geqslant \liminf \int_k^T (A_h \, u^\theta_{h,k} \, , \, u^+_{h,k} - u^\theta_{h,k})dt.$$

Il reste à montrer que le second membre est $\geqslant 0$. Posons

$$Y_{h,k} = \left| \int_0^{T-k} (A_h \, u^\theta_{h,k} \, , \, u^+_{h,k} - u^\theta_{h,k})dt \right|$$

où (cf.(1.82)(1.83)(1.84)(1.60)(1.96))

$$u^\theta_{h,k}(t) = \theta^- \, u^-_{h,k}(t) + \theta^* \, u^*_{h,k}(t) + \theta^+ \, u^+_{h,k}(t) \ .$$

Alors

$$Y_{h,k} = \left| \sum_{i=0}^{N-1} k \, (A_h \, u_h^{i+\theta} \, , \, u_h^{i+1} - u_h^{i+\theta})_h \right|$$

---

(1) On rappelle que l'égalité dans un espace de Lebesgue implique l'égalité p.p. cf. ZAANEN [1] Théorème 2, p. 137 pour l'obtention de (1.110) à partir de (1.109).

et

$$(1.113)\qquad u_{h,k} \leqslant \sqrt{k}\, S(h)\ \Big(\sum_{i=0}^{N} k \| u_h^{i+\theta} \|_h^2\Big)^{\frac{1}{2}} \Big(\sum_{i=0}^{N} | u_h^{i+1} - u_h^{i+\theta} |^2\Big)^{\frac{1}{2}} .$$

Mais

$$u_h^{i+1} = (\theta^- + \theta^* + \theta^+) u_h^{i+1} ,$$

d'où

$$(1.114)\qquad \Big(\sum_{i=0}^{N} | u_h^{i+1} - u_i^{i+\theta} |^2\Big)^{\frac{1}{2}} \leqslant \Big(\sum_{i=0}^{N} (\theta^-)^2 | u_h^{i+1} - u_h^{i} |^2 + \sum_{i=0}^{N} (\theta^*)^2 | u_h^{i+1} -$$

$$- u_h^{i+\frac{1}{2}} |^2 \leqslant {}^{(1)}\ \theta^- \Big(\sum_{i=0}^{N} | u_h^{i+1} - u_h^{i} |^2\Big)^{\frac{1}{2}} + \theta^* \Big(\sum_{i=0}^{N} | u_h^{i+1} - u_h^{i+\frac{1}{2}} |^2\Big)^{\frac{1}{2}} .$$

Donc d'après la condition de stabilité (1.81) et les résultats (1.87) on déduit de (1.113), compte tenu de (1.114), que

$$Y_{h,k} \longrightarrow 0 \quad \text{lorsque} \quad h \ \text{et} \ k \to 0 \ \text{et} \ kS(h)^2 \to 0 .$$

Donc u est solution faible de l'inéquation (1.2) (avec $j=0$). ∎

1.1 4. Convergence forte du schéma III

On raisonne comme aux § 1.10 et 1.13. Il faut montrer que $X_{h,k}$ donné par (1.42) tend vers 0. D'après le Lemme 1.2 , en prenant $v_{h,k} = \tilde{u}_{h,k}$ dans l'inéquation (1.34) (avec j=0 et compte tenu de la Remarque 1.2), (1.34) étant écrit avec $A_h u^{\theta}_{h,k}$, on a

$$\int_k^T \big(\bar{\delta}_k \tilde{u}_{h,k}(t) + A_h(\theta \bar{u}_{h,k}(t-k) + \theta^* u^*_{h,k}(t-k) + \theta^+ u_{h,k}(t)) - f_h , \tilde{u}_{h,k} -$$

$$- u_{h,k}(t)\big)_h dt \ \geqslant\ 0 .$$

On remplace alors $A_h u^{\theta}_{h,k}(t)$ par

$$A_h \left[u_{h,k}(t) + \theta^-(u_{h,k}(t-k) - u_{h,k}(t)) + \theta^*(u^*_{h,k}(t-k) - u_{h,k}(t))\right] .$$

On peut alors passer à la limite en utilisant la même technique de majoration que pour la convergence faible (voir(1.113) et (1.114)). En définitive

$$X_{h,k} \to 0$$

---

(1) $\quad (|a| + |b|)^{\frac{1}{2}} \leqslant |a|^{\frac{1}{2}} + |b|^{\frac{1}{2}} .$

et d'après la coercivité, on obtient

$$p_h\, u_{h,k} \longrightarrow \sigma u \quad \text{dans} \quad L^2(0,T;F) \text{ fort.} \qquad \blacksquare$$

Pour résumer nous rassemblons les résultats concernant la stabilité du schéma III dans le tableau suivant :

Stabilité.

Condition de stabilité $\boxed{1-k\,S(h)^2 \geqslant \beta > 0}$

$\theta^- < \theta^* + \theta^+$	$p_h\, u^{\theta}_{h,k}$ borné inconditionnellement dans $L^2(0,T;F)$ $q_h\, u_{h,k}$ borné inconditionnellement dans $L^\infty(0,T;H)$ $\left.\begin{array}{l} \sum_{i=0}^{N} \lvert u^{i+1} - u^i \rvert^2 \leqslant C \\ \sum_{i=0}^{N} \lvert u^{i+1} - u^i \rvert^2 \leqslant C \\ \sum_{i=0}^{N} \lvert u^{i+1} - u^{i+\frac{1}{2}} \rvert^2 \leqslant C \end{array}\right\}$ inconditionnellement
$\theta^-, \theta^*, \theta^+ \geqslant 0$ quelconques dans les autres cas	sous réserve de la condition de stabilité, on a : $p_h\, u_{h,k}$, $p_h\, u^{*}_{h,k}$, $p_h\, u^{\theta}_{h,k}$ sont bornés dans $L^2(0,T;F)$ $q_h\, u_{h,k}$, $q_h\, u^{*}_{h,k}$, $q_h\, u^{\theta}_{h,k}$ sont bornés dans $L^\infty(0,T;H)$ et : $\sum_{i=0}^{N} \beta \lvert u^{i+1} - u^{i+\frac{1}{2}} \rvert^2 \leqslant C$ $\sum_{i=0}^{N} \beta \lvert u^{i+\frac{1}{2}} - u^{i} \rvert^2 \leqslant C$ $\sum_{i=0}^{N} \beta \lvert u^{i+1} - u^{i} \rvert^2 \leqslant C$

Remarque 1.4 (cas ou $K_h = V_h$).

Lorsque $K_h = V_h$ on a $u_h^{i+1} = P_{V_h}\left[u_h^{i+\frac{1}{2}}\right]$, donc $u_h^{i+1} = u_h^{i+\frac{1}{2}}$ et le schéma (1.57)(1.58) se réduit, en posant $\theta = 1 - \theta^-$, à

$$\frac{u_h^{i+1} - u_h^i}{k} + A_h\left[u_h^i + \theta(u_h^{i+1} - u_h^i)\right] - f_h = 0 .$$

On montre alors, dans le cas ou $\theta \geqslant \frac{1}{2}$ (cf. RAVIART [3] , TEMAM [2]) que lorsque $k,h \to 0$,

$$\left|\begin{array}{l} p_h\, u_{h,k} \longrightarrow \sigma u \quad \text{dans } L^2(0,T;F) \text{ fort,} \\ q_h\, u_{h,k} \longrightarrow u \quad \text{dans } H \text{ fort, p.p. en } t \in [0,T] , \end{array}\right.$$

ces convergences étant inconditionnelles (c'est à dire que la condition $kS(h)^2 \to 0$ n'est pas requise). ■

Remarque 1.5 (cas ou $K_h \neq V_h$) .

Il ne semble pas que l'on puisse démontrer, sauf hypothèses particulières que le schéma III soit inconditionnellement stable et convergent si $\theta^* \geqslant \frac{1}{2}$ (ce qui est le cas des équations). Ceci pourrait s'expliquer par le fait que la projection se fait au sens de la norme $(,)_h$; on garde donc la stabilité dans $L^2(0,T;H)$ mais, à priori, cela n'entraîne en aucune façon que la stabilité dans $L^2(0,T;V)$ soit conservée. Les schémas III ont cependant l'avantage de "découpler" les difficultés numériques mais théoriquement ils ne semblent pas plus intéressants qu'un schéma totalement explicite. ■ On a le

Théorème 1.3.

Avec θ^- , θ^* , $\theta^+ \geqslant 0$, $\theta^- + \theta^* + \theta^+ = 1$, si $h,k \to 0$

de façon que

$$kS(h)^2 \to 0 \qquad \text{si } \theta^+ \neq 1 \tag{1.115}$$

alors les solutions $u_{h,k}$ du schéma III vérifient :

$$p_h\, u_{h,k} \longrightarrow \sigma u \text{ dans } L^2(0,T;F) \text{ faible,} \tag{1.116}$$

$$q_h\, u_{h,k} \longrightarrow u \text{ dans } L^\infty(0,T;H) \text{ faible étoile.} \tag{1.117}$$

où u est la solution unique de l'inéquation (1,2).

Lorsque cette solution est solution forte unique de (1.1) on a

$$p_h u_{h,k} \longrightarrow u \quad \text{dans } L^2(0,T;F) \text{ fort.}$$

L'hypothèse (1.115) est sans objet si

$$\theta^+ = 1$$

ou bien si $K_h = V_h$, $K = V$, (alors $u^{i+1} = u^{i+\frac{1}{2}}$) ,

$$\theta^+ = 0 \quad , \quad \theta^* \geqslant \theta^- \quad . \quad \blacksquare$$

Remarque 1.5.

Avec $K_h = V_h$, $K=V$, on a la propriété classique des schémas semi-implicites(dits de Crank-Nicholson) qui sont inconditionnellement stables et inconditionnellement convergents. ■

Dans tout ce qui précède (et dans le cas général $K_h \neq V_h$) nous n'avons obtenu de convergence inconditionnelle que dans le cas où $\theta^+ = 1$, c'est à dire pour les schémas totalement implicites. On remarquera que du point de vue de la résolution numérique les schémas semi-implicites posent les mêmes difficultés que les schémas totalement implicites. ■

2. VERIFICATION DES HYPOTHESES

Nous allons vérifier les hypothèses fondamentales introduites dans la première partie de ce chapitre. Nous commencerons par rappeler 2 lemmes classiques en approximation (cf. LIONS [1] , CEA [1], RAVIART [1] [2] [3]) Soit g une fonction intégrable. Nous posons, (k étant appelé à tendre vers 0)

(2.0) $$\tau_k g(t) = \frac{1}{2k} \int_{s-k}^{s+k} g(t)dt \ , \qquad t \in]\, s-k,\ s+k[\ .$$

Lemme 2.1.

On a $\tau_k g(t) \to g(t)$, pour presque tout t. ■ (1)

Lemme 2.2.

$\forall g \in L^p(0,T;X)$, $1 \leqslant p \leqslant \infty$, X un espace de Banach, alors

$$\tau_k g \to g \text{ dans } L^p(0,T;X) \text{ fort,}$$

$$\|\tau_k g\|_{L^p(0,T;X)} \leqslant \|g\|_{L^p(0,T;X)} \ . \ \blacksquare \ (1)$$

2.1. Hypothèses sur $\mathcal{H}$

Le lemme qui suit montre que l'hypothèse de consistance (1.10) est vérifiée.

---

(1) Ces deux lemmes sont encore valables avec τ_k défini par :

$$\tau_k g(t) = \frac{1}{k} \int_{s}^{s+k} g(t)dt \qquad , \quad t \in]\, s,\ s+k[\ .$$

Lemme 2.3.

Soient $v \in L^2(0,T;H)$, $\tilde{v} \in L^2(0,T;F)$, et $v_h \in L^2(0,T;K_h)$ tel que :

(2.1)
$$\left| \begin{array}{l} q_h v_h \longrightarrow v \quad \text{dans } L^2(0,T;H) \text{ faible}, \\ p_h v_h \longrightarrow \tilde{v} \quad \text{dans } L^2(0,T;F) \text{ faible}, \end{array} \right.$$

alors, si l'on a l'hypothèse de consistance de K_h ((1.1) Ch.II), on a

$$\left| \begin{array}{l} v \in L^2(0,T;V) \ , \ v(t) \in K \ \text{p.p. en } t \in [0,T] \ , \\ \tilde{v} = \sigma v \ , \ \text{p.p. en } t \in [0,T]. \end{array} \right. \quad \blacksquare$$

Ce Lemme reste valable pour tous les espaces L^p , $p \geqslant 2$. $\blacksquare$

Démonstration

D'après 2.1. on a $\lim q_h v_h = v$ dans $L^2(0,T;H)$ faible donc $\lim q_h v_h = v$, p.p. en $t \in [0,T]$ dans H faible, de même $\lim p_h v_h = \tilde{v}$, p.p. en $t \in [0,T]$ dans F faible. On a alors par linéarité des opérateurs et pour presque tout $t \in [0,T]$

$$\tau_k \lim_{h \to 0} q_h v_h = \lim_{h \to 0} q_h \tau_k v_h = \tau_k v \quad \text{dans H faible},$$

$$\tau_k \lim_{h \to 0} p_h v_h = \lim_{h \to 0} p_h \tau_k v_h = \tau_k \tilde{v} \quad \text{dans F faible}.$$

D'après la convexité de K_h et par linéarité de l'intégrale on a $\tau_k \tilde{v}_h \in K_h$. D'après l'hypothèse de consistance ((1.1) Ch.II), il vient

$$\tau_k \tilde{v} \in \sigma K,$$

(2.3)
$$\tau_k \tilde{v} = \sigma \bar{v}_k \ , \ \bar{v}_k \in K \ ,$$

et comme $\pi \sigma = I$ on a

$$\pi \tau_k \tilde{v} = \bar{v}_k = \pi \lim_{h \to 0} p_h \tau_h v_h =$$

$$= \lim_{h \to 0} \pi p_h \tau_k v_h \ .$$

Mais comme par définition (cf. § 1.1. Ch. IV) on a $\pi p_h = q_h$, il vient

(2.4)
$$\bar{v}_k = \lim_{h \to 0} q_h \tau_k v_h = \tau_k v \quad \text{et l'on a}$$
$$\tau_k \tilde{v} = \tau_k v$$

soit

(2.5)
$$\tau_k \tilde{v} = \sigma \tau_k v \quad \text{et en particulier} \quad \tau_k v \in V \ .$$

Il reste maintenant à passer à la limite avec $k \to 0$. On remarque d'abord que $\tau_k \tilde{v} \in L^2(0,T;F)$ et, d'après le Lemme 2.2 , $\tau_k \tilde{v} \to \tilde{v}$ dans $L^2(0,T;F)$ fort. Comme σ est un isomorphisme on déduit que $\tau_k v$ est suite de Cauchy dans $L^2(0,T;V)$. Donc $\tau_k v$ converge vers $\hat{v}$ dans $L^2(0,T;V)$. D'autre par comme $\tau_k v \to v$ dans

$L^2(0,T;H)$ fort (d'après le Lemme 2.3) on déduit que $\hat{v} = v$ p.p. et donc que

$$v \in L^2(0,T;V) .$$

Mais $\tau_k v \in K$ (d'après (2.3) et(2.4)) et comme

(2.6)
$$\tau_k v \to v \quad \text{p.p. } t \in [0,T]$$

on déduit, puisque K est un convexe fermé, que $v \in K$, p.p. en $t \in [0,T]$. Il reste à vérifier que $\tilde{v} = \sigma v$, p.p. en $t \in [0,T]$. D'après (2.5) on a $\tau_k \tilde{v} = \sigma \tau_k v$, pour presque tout $t \in [0,T]$

d'où
$$\lim_{h \to 0} \tau_k \tilde{v} = \tilde{v} = \sigma \lim_{k \to 0} \tau_k v = \sigma v .$$

Ce qui établit le lemme. ∎

On notera que $L^\infty(0,T;X) \subset \ldots \subset L^2(0,T;X) \subset L^1(0,T;X)$ et que la topologie d'un espace L^r avec $r > p$ est plus fine que la topologie de L^p. ∎

Nous allons montrer que les hypothèses de convergence sur K_h (1.11)(1.12)(1.13) sont vérifiées lorsque la famille K_h est convergente au sens "stationnaire" (cf.(1.2) Ch.II). Tout d'abord (1.13) découle directement de (1.2) Ch.II. Le lemme suivant a pour objet de montrer que (1.11) et (1.12) résultent aussi de (1.2)Ch.II.

Lemme 2.4.
Sous l'hypothèse de convergence (1.2)Ch.II, on a

$$\forall\, v \in L^2(0,T;K) \ , \exists\ v_{h,k} \in L^2(0,T;K_h)$$

tel que

(2.7) $\quad q_h v_{h,k} \longrightarrow v$ dans $L^2(0,T;H)$ fort,

(2.8) $\quad p_h v_{h,k} \longrightarrow \sigma v$ dans $L^2(0,T;F)$ fort.

De plus, si $v' \in L^2(0,T;V')$, alors

(2.9) $\quad q_h \delta_k v_{h,k} \longrightarrow v'$ dans $L^2(0,T;V')$ fort. ∎

Démonstration.
Soit $v \in L^2(0,T;V)$, $v \in K$ p.p. en $t \in [0,T]$.

Posons ici

(2.10)
$$\tau_k v = \sum_{i=0}^{N} \frac{1}{k} \int_{ik}^{(i+1)k} v(t)dt \ . \ \chi_k^i$$

où χ_k^i est la fonction caractéristique de l'intervalle $[ik, (i+1)k[$.

Nous allons montrer qu'il existe une suite de fonctions $v_{h,k}$ telle que

$$p_h\, v_{h,k} \longrightarrow \sigma v \quad \text{dans } L^2(0,T;F) \text{ fort}$$

ou, ce qui revient au même, telle que

$$\| \sigma v - p_h\, v_{h,k} \|_{L^2(0,T;F)} \leqslant 0(h,k)$$

ou encore, puisque $\sigma v - p_h\, v_{h,k} = \sigma v - \sigma \tau_k v + \sigma \tau_k v - p_h\, v_{h,k}$ et d'après le résultat du Lemme 2.2 valable aussi pour l'opérateur τ_k que nous considérons ici, il suffira de montrer que

$$\| \sigma \tau_k v - p_h\, v_{h,k} \|_{L^2(0,T;F)} \leqslant 0(h,k) \,. \tag{2.11}$$

Comme $\tau_k v \in K$ cela résulte sans difficulté de l'hypothèse de convergence (1.2) Ch.I

La convergence $q_h\, v_{h,k} \longrightarrow v$ dans $L^2(0,T;H)$ se déduit de la définition de q_h .

Supposons maintenant que $v' \in L^2(0,T;V')$. Nous considérons des fonctions Λ_k^i telles que ci-dessous.

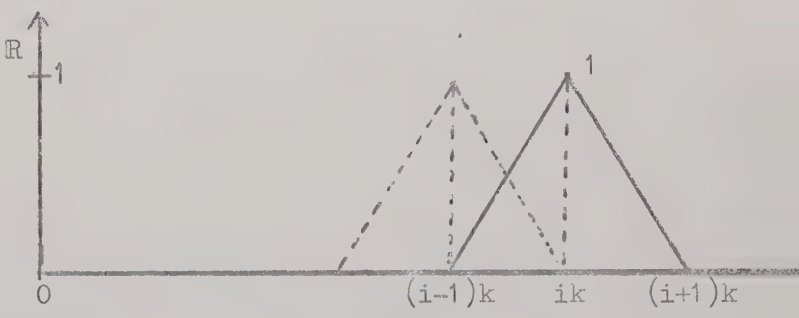
Nous posons

$$v^i = \frac{1}{k} \int_{(i-1)k}^{ik} v(t)dt$$

et

$$\tau_k v = \sum_{i=1}^{N+1} v^i .\ \Lambda_k^i \,. \tag{2.12}$$

On prolongera v par 0 en dehors de $[0,T]$.



Nous allons d'abord montrer que :

$$\left| \begin{array}{l} (\tau_k v)' \in L^2(0,T;V') \text{ ce qui est évident et que} \\ (\tau_k v)' \longrightarrow v' \text{ dans } L^2(0,T;V') \text{ fort.} \end{array} \right. \tag{2.13}$$

Evaluons d'abord

$$(\tau_k v(t))' \text{ que nous noterons } \tau_k' v \,.$$

Pour $t \in [ik,(i+1)k[$ on a

$$\tau_k' v(t) = \frac{1}{k^2} \int_{ik}^{(i+1)k} (v(t) - v(t-k))dt$$

or

$$u(t) - u(t-k) = \int_{t-k}^{t} v(s)ds$$

d'où

$$\tau'_k v(t) = \frac{1}{k^2} \int_{ik}^{(i+1)k} dt \int_{t-k}^{t} v'(s)ds$$

soit, en échangeant l'ordre des intégrales, il vient

$$\tau'_k v(t) = \frac{1}{k} \int_{(i-1)k}^{(i+1)k} v' \Lambda_k^i \, ds \ .$$

Nous allons voir que $\tau'_k v \to v'$ dans $L^2(0,T;V')$ fort. Pour cela il est équivalen de voir (nous redéfinissons τ_k) que

$$\tilde{\tau}_k v = \sum_{i=0}^{N} \frac{1}{k} \int_{(i-1)k}^{(i+1)k} v \Lambda_k^i \, dt \ . \ \chi_k^i \to v \quad \text{dans } L^2(0,T,V') \text{ fort} \ .$$

Avant de la prouver nous remarquerons que

$$\| \tilde{\tau}_k v - \tilde{\tau}_k u \|_{L^2(0,T;V')} \leq \| v-u \|_{L^2(0,T;V')}$$

puisque

$$\frac{1}{k} \int_{(i-1)k}^{(i+1)k} \Lambda_k^i = 1 \ .$$

Soit alors $\varphi_\varepsilon \in \mathcal{C}^0(0,T;V')$, espace dense dans $L^2(0,T,V')$ tel que

$$\| v - \varphi_\varepsilon \|_{L^2(0,T;V')} \leq \frac{\varepsilon}{2} \ .$$

Alors (les normes étant prises dans $L^2(0,T;V')$)

$$\| \tilde{\tau}_k v - v \| \leq \| \tilde{\tau}_k v - \tilde{\tau}_k \varphi_\varepsilon \| + \| \tilde{\tau}_k \varphi_\varepsilon - \varphi_\varepsilon \| + \| \varphi_\varepsilon - v \| \ .$$

D'où

$$\| \tilde{\tau}_k v - v \| \leq \varepsilon + \| \tilde{\tau}_k \varphi_\varepsilon - \varphi_\varepsilon \| .$$

Comme φ_ε est une fonction continue il est clair que $\tau_k \varphi_\varepsilon \to \varphi_\varepsilon$ lorsque $k \to 0$
D'où

$$\| \tilde{\tau}_k v - v \| \leq 0(k) \qquad (\text{i.e.} \to 0 \text{ lorsque } k \to 0) \ . \tag{2.14}$$

Ce résultat prouve (2.13). Il reste à prouver (2.11). Soit $v_{h,k}$ une fonction pour l'instant quelconque. On a (les normes étant toujours prises dans $L^2(0,T,V')$)

$$\| q_h \delta_k v_{h,k} - v' \| \leq \| q_h \delta_k v_{h,k} - \tilde{\tau}_k v' \| + \| \tilde{\tau}_k v' - v' \| \ .$$

Il reste à montrer que le 1er terme du second membre peut être rendu aussi petit que l'on veut avec une suite $v_{h,k}$ convenable (qui doit vérifier aussi (2.11)), i

$$\| q_h \delta_k v_{h,k} - \tilde{\tau}_k v' \| \leq 0(k,\varepsilon) \ .$$

Or $\tilde{\tau}_k v' = \tau'_k v$ et pour $t \in [ik,(i+1)k[$ on a

$$\tau'_k v = \frac{v^{i+1} - v^i}{k} .$$

Donc

$$\| q_h \delta_k v_{h,k} - \tau'_k v \| = \frac{\| q_h v^{i+1}_{h,k} - v^{i+1} \|}{k} + \frac{\| q_h v^i_{h,k} - v^i \|}{k} .$$

En prenant alors $v_{h,k}$ tel que

$$\| q_h v^i_{h,k} - v^i \| \leq 0(\sqrt{k}, h) , \quad i = 0,1...,N .$$

On prouve (2.9) et l'on assure en même temps (2.11.). Le Lemme 2.4 est donc prouvé.

2.2. Hypothèses sur $a(t;u,v)$

Nous allons vérifier les hypothèses introduites pour l'approximation de $a(t;u,v)$ pour une famille particulière de formes $a_{h,k}(t; u_{h,k}, v_{h,k})$. Tout d'abord on suppose qu'il est possible de définir une extension $\tilde{a}(t; \sigma u, \sigma v)$ de $a(t;u,v)$ à F, qui conserve les mêmes propriétés de bilinéarité, de continuité et de coercivité. On suppose que $\|v_h\|_h = \|p_h v_h\|_F$. On définit, pour $i = 0,...,1$,

(2.15) $$\tilde{a}^i_k(u,v) = \frac{1}{k} \int_{ik}^{(i+1)k} \tilde{a}(t;u,v)dt \quad , \quad u,v \in F ,$$

(2.16) $$\tilde{a}^i_{h,k}(u_h , v_h) = \tilde{a}^i_k(p_h u_h , p_h v_h) ,$$

et

(2.17) $$a_{h,k}(t; u_{h,k} , v_{h,k}) = \sum_{i=0}^{N} \tilde{a}^i_{h,k} (u^i_h , v^i_h).k$$

avec

$$u_{h,k} = \sum_{i=0}^{N} u^i_h \chi^i_k \quad , v_{h,k} = \sum_{i=0}^{N} v^i_h \chi^i_k .$$

Alors

$$|a_{h,k}(t ; u_h , v_h)| \leq C \| u_h \|_h \| v_h \|_h \quad , \forall u_h \in V_h , \forall v_h \in V_h ,$$

$$a_{h,k}(t ; u_h , u_h) \geq \alpha \|u_h\|^2_h \quad , \quad \forall u_h \in V_h ,$$

donc les propriétés (1.15)(1.16) sont vérifiées. Nous allons montrer que si pour tout $t \in [0,T]$, $\{ a_{h,k}(t ; u_h , v_h)\}$ est consistante et convergente au sens (1.5 (1.6) Ch.II alors

$$\{ a_{h,k}(t ; u_h , v_h)\}$$

est consistante et convergente au sens (1.17)(1.18) Ch.IV.

Lemme 2.5 (consistance). Etant donné $v_{h,k}(t) \in K_h$, $\forall t \in [0,T]$ tel que

(2.18) $$p_h\, v_{h,k} \longrightarrow \sigma v \quad \text{dans } L^2(0,T;F) \text{ faible}$$

alors

(2.19) $$\liminf \int_0^T a_{h,k}(t\,;\, v_{h,k}\,,\, v_{h,k})dt \geqslant \int_0^T a(t\,;\, v,v)dt.$$

Démonstration

Notons d'abord qu'il suffit de prouver ce lemme pour $v_{h,k} \in V_h$ (au lieu de K_h). Un premier point est de voir que les intégrales sont bien définies. En effet, pour l'intégrale de droite, cela résulte des hypothèses (1.5). Pour l'intégrale de gauche on a

$$|a_{h,k}(t,\, v_{h,k}\,,\, v_{h,k})| \leqslant C \;\|v_{h,k}(t)\|_h^2 \quad , \quad \forall t \in [0,T],$$

mais $\|p_h\, v_{h,k}(t)\| = \|v_{h,k}(t)\|_h$ et comme $p_h\, v_{h,k} \longrightarrow \sigma v$ dans $L^2(0,T;F)$ faible il résulte que $\int_0^T \|v_{h,k}(t)\|^2 dt < +\infty$. D'où l'on déduit, en utilisant le Théorème de Lebesgue que l'intégrale de gauche existe. Montrons maintenant (2.1 On a, d'après (2.17) et (2.15),

(2.20) $$\int_0^T a_{h,k}(t\,;\, v_{h,k}\,,\, v_{h,k})dt = \int_0^T \tilde{a}(t\,;\, p_h\, v_{h,k}\,,\, p_h\, v_{h,k})\, dt.$$

Mais d'après les Lemmes 2.7 et 2.8 Ch.III la fonction ainsi définie est semi-continue-inférieurement, et comme elle est convexe, elle est faiblement s.c.i., d'où le résultat (2.19). ∎

Le lemme suivant élimine l'hypothèse de convergence (1.18).

Lemme 2.6 (convergence). Etant donnés $v \in L^2(0,T;K)$, $w \in L^2(0,T;K)$ et

$v_{h,k}(t) \in K_h$, $w_{h,k}(t) \in K_h$, $t \in [0,T]$ tels que

$p_h\, v_{h,k} \longrightarrow \sigma v$ dans $L^2(0,T;F)$ faible,

$p_h\, w_{h,k} \longrightarrow \sigma w$ dans $L^2(0,T;F)$ fort,

alors

(2.21) $$\int_0^T a_{h,k}(t\,;\, v_{h,k}\,,\, w_{h,k})dt \longrightarrow \int_0^T a(t\,;\, v,w)dt,$$

(2.22) $$\int_0^T a_{h,k}(t\,;\, w_{h,k}\,,\, v_{h,k})dt \longrightarrow \int_0^T a(t\,;\, v,w)dt. \quad ∎$$

Démonstration

Comme précédemment il suffit de prouver ce lemme sur V_h. Cela résulte sans difficulté de (2.20) et des propriétés de $\tilde{a}(t;\tilde{v},\tilde{w})$ (bilinéarité et continuité et le fait qu'une suite qui converge faiblement reste bornée). ∎

2.3. Hypothèses sur $j(t,v)$

On suppose [1] qu'il est possible d'étendre $j(t,v)$ à F par une fonction $\tilde{j}(t,\tilde{v})$, $\tilde{v} \in F$, qui garde les mêmes propriétés que $j(t,v)$ (semi-continuité inférieure, convexité, majoré, $\not\equiv \pm\infty$ sur σK) et nous supposerons que $\tilde{j}(t, p_h v_h) \not\equiv \pm\infty, \forall v_h \in$

On définit

$$\tilde{j}_k^i(v) = \frac{1}{k} \int_{ik}^{(i+1)k} \tilde{j}(t,v)dt \quad , \quad v \in F \ ,$$

$$\tilde{j}_{h,k}^{\ i}(v) = \tilde{j}_k^i(p_h v_h) \ ,$$

$$j_{h,k}(t,v_{h,k}) = \sum_{i=0}^{N} \tilde{j}_{h,k}^i(v_h^i)k \quad , \quad \text{avec} \quad v_{h,k} = \sum_{i=0}^{N} v_h^i \ \chi_k^i \ .$$

Il est alors immédiat de vérifier que $j_{h,k}(t,v_{h,k})$ est s.c.i., convexe, majorée, $\not\equiv \pm\infty$ sur K_h). De plus

$$\int_0^T j_{h,k}(t,v_{h,k}) = \int_0^T \tilde{j}(t,p_h v_{h,k}) \ .$$

Grâce aux Lemmes 2.7 et 2.8 Ch. III on obtient sans difficulté le

Lemme 2.7 (consistance). Si l'on a $v_{h,k}(t) \in K_h$, $\forall t \in [0,T]$ tel que

$$p_h v_{h,k}(t) \longrightarrow \sigma v \quad \text{dans} \quad L^2(0,T;F) \quad \text{faible}$$

alors

$$\liminf \int_0^T j_{h,k}(t,v_{h,k})dt \geq \int_0^T j(t,v)dt \ . \qquad (2.23)$$

Pour assurer la convergence, il nous faut supposer en plus que $\tilde{j}(t,\tilde{v})$ est continu sur F . On a alors directement le

Lemme 2.8 (convergence). Si l'on a $v_{h,k}(t) \in K_h$, $\forall t \in [0,T]$ tel que

$$p_h v_{h,k} \longrightarrow \sigma v \quad \text{dans} \quad L^2(0,T;F) \quad \text{fort}$$

alors

$$\int_0^T j_{h,k}(t , v_{h,k})dt \to \int_0^T j(t,v)dt \ . \qquad (2.24)$$

---

(1) cette condition n'est cependant pas toujours réalisable.

2.4. Hypothèses sur f

On définit, de façon générale,

$$f^i_{h,k} \quad \text{par} \quad (f^i_{h,k}\ ,\ v_h)_h = \frac{1}{k} \int_{ik}^{(i+1)k} \langle f(t),\ p_h v_h \rangle_{(F',F)}\, dt,$$

$$f_{h,k} = \sum_{i=0}^{N} f^i_{h,k} \quad \chi^i_k \ .$$

Si $f \in L^2(0,T;H)$ on peut prendre

$$f^i_{h,k} \quad \text{défini par} \quad (f^i_{h,k}\ ,\ v_h)_h = (\frac{1}{k} \int_{ik}^{(i+1)k} f(t)dt,\ q_h v_h).$$

On montre alors, et cela est classique[1] que les hypothèses (1.26) et (1.27) sont vérifiées. ∎

(1) cf. LIONS [1] , RAVIART [1][2][3] , NEDELEC [1] .

CHAPITRE V

EXISTENCE ET UNICITE POUR LES PROBLEMES HYPERBOLIQUES OU BIEN POSES AU SENS DE PETROWSKI

1. FORMULATION ABSTRAITE

On se donne :

(1.1) V , espace de Hilbert ; produit scalaire $((,))$; norme $\|.\|$

(1.2) V', dual de V ; produit scalaire canonique $((,))_*$; norme $\|.\|_*$

(1.3) H , espace de Hilbert ; produit scalaire $(,)$; norme $|.|$

H', dual de H ; H' identifié à H , avec (comme en (1.3.) Ch.III)

$$V \subset H \subset V'$$

(1.4) K , convexe fermé non vide de V

(1.5) $t \to a(t;u,v)$, fonction mesurable sur $[0,T]$, $T < +\infty$, ceci $\forall u,v \in V$

$t \to a(t;u,v) \in \mathcal{C}^1(0,T)$, $\forall u,v \in V$,

et, pour presque tout $t \in [0,T]$,

(1.5.a) $a'(t;v,v) \leqslant 0$, $\forall v \in V$,

$a(t;u,v)$ forme bilinéaire, monotone, symétrique, uniformément continue en t , i.e.

$|a(t;u,v)| \leqslant C \|u\| \cdot \|v\|$, $\forall u,v \in V$, C constante indépendante de t,u,v.

(1.6) $t \to j(t,v)$, mesurable, $\forall v \in V$, avec, pour presque tout $t \in [0,T]$,

$|j(t,v)| \leqslant g(t)$, fonction intégrable, ceci $\forall v \in K$,

$j(t,v)$ convexe, s.c.i. sur K

(1.7) $f \in L^2(0,T;V)$

(1.8) $u_0 \in V$, $u_1 \in K$

On définit les espaces

$$X(0,T) = \{ v \mid v, v' \in L^2(0,T;V) \} , \tag{1.9}$$

$$X_o(0,T) = \{ v \mid v \in X(0,T),\ v(0) = u_o \} , \tag{1.10}$$

$$Y(0,T) = \{ v \mid v \in X(0,T),\ v'' \in L^2(0,T;V') \} , \tag{1.11}$$

$$Y_o(0,T) = \{ v \mid v \in Y(0,T),\ v(0) = u_o ,\ v'(0) = u_1 \} , \tag{1.12}$$

et les convexes

$$\mathcal{K} = \{ v \mid v \in Y_o(0,T),\ v'(t) \in K ,\ t \in [0,T] \} , \tag{1.13}$$

$$\mathcal{K}_f = \{ v \mid v \in X_o(0,T),\ v'(t) \in K ,\ t \in [0,T] \} . \tag{1.14}$$

On définit l'opérateur

$$A(t) \in \mathcal{L}(V,V') ,\ (A(t)u,v) = a(t;u,v) ,\ \forall u,v \in V . \tag{1.15}$$

Nous allons étudier le problème (1) :

$$\left| \begin{array}{l} \text{trouver } u \in \mathcal{K} \text{ tel que } \forall v \in \mathcal{K} \text{ on ait} \\ \\ L(u,v) = \displaystyle\int_0^T (u'' + Au - f ,\ v' - u')dt + \int_0^T [\, j(v') - j(u') \,]\, dt \geqslant 0 . \end{array} \right. \tag{1.16}$$

On définit, avec $R = (R_1 , R_2 , R_3) \in \bar{\mathbb{R}}_+^3$ (2),

$$\mathcal{K}_R = \{ v \mid v \in \mathcal{K} ,\ \|v\|_{L^2(0,T;V)} \leqslant R_1 ,\ \|v'\|_{L^2(0,T;V)} \leqslant R_2 ,\ \|v''\|_{L^2(0,T;V')} \leqslant R_3 \} \tag{1.17}$$

et l'on munit les espaces $X(0,T)$ et $Y(0,T)$ des produits scalaires et normes canoniques :

$$((u,v))_{X(0,T)} = ((u,v))_{L^2(0,T;V)} + ((u',v'))_{L^2(0,T;V)} , \tag{1.18}$$

$$((u,v))_{Y(0,T)} = ((u,v))_{X(0,T)} + ((u'',v''))_{L^2(0,T;V')} . \tag{1.19}$$

2. RESULTATS PRELIMINAIRES

On donne sans démonstration un certain nombre de lemmes classiques et quelques autres dont la démonstration n'offre pas de difficulté.

Lemme 2.1.

Muni du produit scalaire $((u,v))_{Y(0,T)}$ (donné par (1.19)) l'espace $Y(0,T)$ est un espace de Hilbert. De plus, si $v \in Y(0,T)$, en modifiant éventuellement v sur un ensemble de mesure nulle, v est continu de $[0,T]$ dans V et V' est

---

(1) On omet les paramètres t pour alléger les notations.

(2) $\bar{\mathbb{R}}_+^3 = [\, \mathbb{R}_+ \cup \{+\infty\} \,]^3$.

continu de $[0,T]$ dans H. ∎

Lemme 2.2.

L'intégrale $\int_0^T (v'',v')dt$ est une fonction convexe sur l'espace

$$Y_1(0,T) = \{ v \mid v \in Y(0,T),\ v'(0) = u_1 \} .$$ ∎

Lemme 2.3.

L'intégrale précédente est une fonction continue sur $Y_1(0,T)$. ∎

Lemme 2.4.

Soit $A(t) \in \mathcal{L}(V,V')$ tel que

$$(A(t)v,w) = a(t;v,w), \quad \forall v,w \in V.$$

Si $t \to a(t;v,w) \in \mathcal{C}^1(0,T)$, $\forall v,w$ alors $\frac{d}{dt} a(t;u,v)$ est, $\forall t \in [0,T]$, une forme bilinéaire sur V, continue au sens suivant

$$\left| \frac{d}{dt} a(t;u,v) \right| \leq C \|u\| \|v\| , \quad C \text{ constante indépendante de } t, u, v ,$$

de plus $t \to A(t) : ([0,T] \to \mathcal{L}(V,V')) \in \mathcal{C}^1(0,T)$ et

$$(A'(t)u,v) = \frac{d}{dt} a(t;u,v).$$ ∎

Lemme 2.5.

L'intégrale $\int_0^T (A(t)v,v')dt$ est une fonction convexe sur $X_0(0,T)$ et continue sur $X(0,T)$. ∎

Lemme 2.6.

On a, $\forall v \in X(0,T)$, $A(t)$ étant symétrique,

$$\int_0^T (A(t)v,v')dt = \int_0^T \tfrac{1}{2} \frac{d}{dt} (A(t)v,v)dt - \tfrac{1}{2} \int_0^T (A'v,v)dt , \tag{1.20}$$

$$\int_0^T \tfrac{1}{2} \frac{d}{dt} (A(t)v,v)dt = \tfrac{1}{2}(A(T)v(T),v(T)) - \tfrac{1}{2}(A(0)v(0),v(0)). \tag{1.21}$$ ∎

Lemme 2.7.

Si K n'est pas borné et si

(1.22) $j(t,v)$ est fortement convexe en t, i.e.

$j(t,v)$ est convexe en $v \in K$

$$j\left(t, \frac{v_1 + v_2}{2}\right) - j(v_1) \leq \tfrac{1}{2}[j(v_2) - j(v_1)] - \beta \|v_2 - v_1\|^2 ,$$

ceci $\forall v_1, v_2 \in K$, avec $\beta > 0$, indépendant de t, v_1, v_2,

alors il est possible de définir un convexe fermé borné $\widetilde{K}$ de V tel que les deux problèmes suivants soient équivalents :

$$\left|\begin{array}{l} L(u,v) \geqslant 0 \quad , \quad \forall v \in \mathcal{K}_{R_*} \, , \; u \in \mathcal{K}_R \, , \\ L(u,v) \geqslant 0 \quad , \quad \forall v \in \widetilde{\mathcal{K}}_{R_*} \, , \; u \in \widetilde{\mathcal{K}}_R \, , \end{array}\right.$$

où $\widetilde{\mathcal{K}}_R$ et $\widetilde{\mathcal{K}}_{R_*}$, sont définis comme $\mathcal{K}_R$ et $\mathcal{K}_{R_*}$ en remplaçant K par $\widetilde{K}$, avec R, $R_* \in \mathbb{R}^3_+$. ∎

Remarque 1.1.

La coercivité éventuelle de $A(t)$ n'est pas une notion suffisante pour remplacer (1.22) . Avec la monotonie de $A(t)$ on prouve aisément que $u \in$ borné de $L^\infty(0,T;H)$ mais non pas un borné de $L^2(0,T;V)$. ∎

3. EXISTENCE DANS LE CAS BORNE : SOLUTIONS FORTES

Nous donnons un premier théorème d'existence dans le cas de convexes bornés de $V(0,T$

Théorème 3.1.

Sous les hypothèses (1.1) ...(1.8) avec R, $R_* \in \mathbb{R}^3_+$, les problèmes :

$$(3.1) \qquad \int_0^T (u'' + Au - f, v' - u')dt + \int_0^T [j(v') - j(u')]\,dt \geqslant 0, \quad \forall v \in \mathcal{K}_{R_*} \, , \; u \in \mathcal{K}_R \, ,$$

avec $R_* \leqslant R$,

$$(3.2) \qquad \int_0^T (v'' + Av - f, v' - u')dt + \int_0^T [j(v') - j(u')]\,dt \geqslant 0 \, , \forall \, v \in \mathcal{K}_R \, , \; u \in \mathcal{K}_{R_*} \, ,$$

avec $R_* \leqslant R$,

ont chacun au moins une solution. Avec $R = R_*$, ces problèmes ont le même ensemble de solutions. ∎

Démonstration

On pose

$$(3.3) \qquad L(u,v) = \int_0^T (u'' + Au - f, \, v' - u')dt + \int_0^T [j(v') - j(u')]\,dt \, .$$

Avec les hypothèses faites on vérifie sans difficulté que l'on peut appliquer le Théorème de Ky-Fan-Sion (Théorème 2.1, Ch.I) : il existe $(u_1, u_2) \in \mathcal{K}_R \times \mathcal{K}_R$ tel que

$$(3.4) \qquad L(v, u_2) \leqslant L(u_1, u_2) \leqslant L(u_1, w) \, , \quad \forall (v,w) \in \mathcal{K}_R \times \mathcal{K}_{R_*} .$$

Avec $R_* \leqslant R$, faisant $v = u_2$ dans (3.4) on obtient

$$L(u_1, w) \geqslant 0 \, , \quad \forall \, w \in \mathcal{K}_{R_*} \, , \quad u_1 \in \mathcal{K}_R \, ,$$

et ceci prouve l'existence de u_1 solution de(3.1). Avec $R_* \geq R$, faisant $w = u_2$ dans (3.4) on obtient

$$L(v,u_2) \leq 0 \quad , \quad \forall v \in \mathcal{K}_R \ , \ u \in \mathcal{K}_{R_*}$$

et ceci prouve l'existence de u_2 solution de (3.2) . L'équivalence n'offre pas de difficulté (on remplace $u'' + Au$ dans (3.1) par $u''-v''+v''+Au-Av+Av$ et pour la réciproque on remplace v dans (3.2) par $w = u + \theta(v-u)$, $\theta \in [0,1]$). ∎

4. EXISTENCE DANS LE CAS BORNE : SOLUTIONS FAIBLES

Nous affaiblissons les hypothèses précédentes et démontrons le

Théorème 4.1.

Sous les hypothèses (1.1.)...(1.8.) avec $R \in \mathbb{R}^3_+$ et K borné le problème

$$(4.1) \qquad \int_0^T (v'' + Av - f, v'-u')dt + \int_0^T [j(v') - j(u')]\, dt \geq 0 \ , \quad \forall v \in \mathcal{K}_R \ , \ u \in \mathcal{K}_f$$

a au moins une solution. ∎

Démonstration

On pose

$$(4.2) \qquad L(u,v) = \int_0^T (v'' + Av-f \ , \ v'-u')dt + \int_0^T [j(v') - j(u')]\, dt$$

et l'on vérifie sans difficulté les hypothèses du Théorème de Ky-Fan-Sion (on notera que K borné $\Longrightarrow$ $\mathcal{K}_f$ borné dans $X(0,T)$). Soit alors (u_1, u_2) un point de selle de $L(u,v)$. Comme $\mathcal{K}_f \supset \mathcal{K}_R$, on peut prendre $v=u_2$ dans

$$(4.3) \qquad L(v,u_2) \leq L(u_1 \ , \ u_2) \leq L(u_1,w) \ , \ \forall (v,w) \in \mathcal{K}_f \times \mathcal{K}_R \ ,$$

d'où

$$L(u_1 \ , \ w) \geq 0 \quad , \quad \forall w \in \mathcal{K}_R \ , \ u_1 \in \mathcal{K}_f \ ,$$

ce qui prouve l'existence de u_1 solution de (4.1). ∎

Théorème 4.2.

Sous les hypothèses (1.1)...(1.8) avec K borné, le problème

$$(4.4) \qquad \int_0^T (v'' + Av-f \ , \ v'-u')dt + \int_0^T [j(v')-j(u')]\, dt \geq 0 \ , \quad \forall v \in \mathcal{K} \ , \ u \in \mathcal{K}_f$$

a au moins une solution. ∎

(La démonstration est semblable à celle du Théorème 4.2,Ch.III).

5. PROBLEMES PSEUDO-INSTANTANES

Théorème 5.1.

L'inéquation (3.1.) est équivalente à , $(R \in \bar{R}^3)$,

$$(u'' + Au - f, v' - u') + j(v') - j(u') \geqslant 0 \quad , \quad \forall v \in \mathcal{K}_R \quad , \quad u \in \mathcal{K}_R \, , \tag{5.1}$$

ceci $\forall t \in [0,T]$ sauf peut-être pour un ensemble $X \subset [0,T]$ de mesure nulle.

Démonstration

Avec les hypothèses de mesurabilité on obtient (3.1) en intégrant (5.1).. Pour la réciproque, on opère comme au Théorème 5.1, Ch. III, mais en remplaçant v' dans (3.1)` par

$$u'(t) + \psi(t) \, [w'(t) - u'(t)] \qquad \text{avec } w'(t) \in \mathcal{K}_R \quad .$$

On obtient

$$\int_0^T (u'' + Au - f, w' - u') \, \varphi(t) dt + \int_0^T [j(w') - j(u')] \, \varphi(t) dt \geqslant 0 \, ,$$

$\forall \varphi(t) \geqslant 0$, $\varphi(t) \in \mathcal{D}(0,T)$, d'où (5.1) . ∎

6. PROBLEMES INSTANTANES

Théorème 6.1.

L'inéquation (3.1)` est équivalente à, $(R \in \bar{R}^3_+)$,

$$(u'' + Au - f, v' - u') + j(v') - j(u') \geqslant 0 \quad , \quad \forall v \in K \, , \; u \in \mathcal{K}_R \quad , \quad \text{p.p. sur } [0,T]. \tag{6.1}$$

Démonstration

On opère comme au Théorème 5.1, Ch.III, mais en posant

$$v'(t) = u'(t) + \psi(t) \, [w - u'(t)] \, .$$

Pour la réciproque, on considère une subdivision de $[0,T]$ en N intervalles égaux de longueur k et l'on fait l'hypothèse suivante (cf. Ch.IV),

$$\left| \begin{array}{l} \forall v' \in \mathcal{K}^* = \{ v \mid v \in W(0,T), \; v(0) = u_1 \, , \; v(t) \in K \;\; t \in [0,T] \} \, , \\ \exists \; v_k(t) = \sum_{i=1}^{N} v_i \; \psi_{k,i}(t) \quad \text{avec } v_i \in K, \; \psi_{k,i}(t) \in [0,1] \, , \; \psi_{k,i} \in L^1(0,T), \\ \qquad\qquad \sum_{i=1}^{N} \psi_{k,i}(t) = 1 \, , \; \text{p.p. sur } [0,T] \, , \end{array} \right.$$

tel que

$$\left| \begin{array}{l} v_k \to v' \quad \text{dans } W(0,T) \text{fort}, \\ \limsup \int_0^T \sum_{i=1}^{N} \psi_{k,i}(t) \, j(v_i) dt \leqslant \int_0^T j(w) dt \, . \end{array} \right.$$

On opère alors comme au Théorème 5.1, Ch.III. ∎

7. UNICITE.

On démontre sans difficulté le

Théorème 7.1.

Si $a(t;u,v)$ est t-coercif et /ou si $j(t,v)$ est t-fortement convexe le problème (3.1) a une solution unique. ∎

On rappelle le théorème fondamental (cf. LIONS [5]) que nous démontrons plus loin par des méthodes d'approximation.

Théorème 7.2.

Avec

$$f, f' \in L^2(0,T;H) \; ; \tag{7.1}$$

$$Au_o \in H \; , \tag{7.2}$$

$$u_1 \in K, \tag{7.3}$$

il existe une fonction u unique vérifiant

$$(u'' - Au - f, v-u') \geqslant 0, \quad \forall\, v \in K \; , \; u \in \mathcal{K} \; , \tag{7.4}$$

et

$$u, u' \in L^{\infty}(0,T;V) \; , \tag{7.5}$$

$$u'' \in L^{\infty}(0,T;H) \; . \tag{7.6}$$

∎

Remarque 7.1.

Le résultat 7.4 permet, si K n'est pas borné dans V, de remplacer K par un convexe K_R convenable (R assez grand) qui soit borné dans V et ceci sans rien changer à l'inéquation. ∎

C H A P I T R E VI

APPROXIMATION DES PROBLEMES HYPERBOLIQUES OU BIEN POSES AU SENS DE PETROWSKI

1. THEORIE

ORIENTATION

Nous considérons le problème

$$(1.1)\qquad \int_0^T (u'' + Au - f,\ v' - u')\,dt + \int_0^T [\,j(v') - j(u')\,]\,dt \geqslant 0\ ,\quad \forall\, v \in \mathcal{K}\ ,\ u \in \mathcal{K},$$

dont nous savons qu'il est équivalent à

$$(1.2)\qquad (u'' + Au - f,\ v' - u') + j(v') - j(u') \geqslant 0,\quad t \in [0,T]\ ,\quad \forall\, v \in \mathcal{K}\ ,\ u \in \mathcal{K}\ .$$

Nous allons définir une famille d'inéquations "proches" de (1.2) et dont il est possible de calculer "plus simplement" les solutions. Les solutions de ces inéquations seront cherchées dans des espaces $L^2(0,T;V_h)$, V_h étant de dimension finie. Appelant u_h l'une de ces solutions nous allons montrer que :

$$\begin{aligned} &u_h \longrightarrow u && \text{dans } L^\infty(0,T;V) \text{ faible étoile,}\\ &"u_h' " \longrightarrow u' && \text{dans } L^\infty(0,T;V) \text{ faible étoile,}\\ &"u_h'' " \longrightarrow u'' && \text{dans } L^\infty(0,T;H) \text{ faible étoile,} \end{aligned}$$

où u est solution de l'inéquation (1.1) , montrant par la même occasion que (1.1) possède effectivement une solution sous des hypothèses particulières.

1.1 Cadre fonctionnel.

Nous reprenons le même cadre fonctionnel que celui qui a été défini au N°1.1 du Ch. IV. On se donne une subdivision de $[0,T]$ en $N=N(k)$ intervalles de longueur k, k étant destiné à tendre vers 0. Nous chercherons une "approximation" de $u(t)$ dans la classe des fonctions en escalier sur $[0,T]$ à valeurs dans V_h. Nous noterons comme d'habitude les fonctions approchées par $v_{h,k}$ avec

$$v_{h,k}(t) = \sum_{i=0}^{N} v_h^i \chi_k^i(t), \quad v_h^i \in V_h .$$

On notera

$$\delta v_{h,k}(t) = \frac{v_{h,k}(t+k) - v_{h,k}(t)}{k} ,$$

$$d v_{h,k}(t) = \frac{v_{h,k}(t+k) - v_{h,k}(t-k)}{2k} ,$$

$$\gamma v_{h,k}(t) = \frac{\delta v_{h,k}(t) - \delta v_{h,k}(t-k)}{k} .$$

Pour l'approximation de u_o, u_1, $\mathcal{K}^*$, $a(u,v) = (Au,v)$, $f(t)$ on fait les hypothèses ci-après :

1.2 Approximation de $\mathcal{K}$ (ou de $Y_o(0,T)$ lorsque $\mathcal{K} = Y_o(0,T)$).

On reprend (1.4)(1.13) ch.V pour la définition de K et de $\mathcal{K}$. On considère

(1.3) K_h convexe fermé de V_h.

On se donne deux approximations des valeurs initiales :

(1.4) $u_o \in V$ et $u_1 \in K$; soit $u_{o,h} \in V_h$ tel que $p_h u_{o,h} \to \sigma u_o$ dans F fort et $\|u_{o,h}\|_h \leq C$,

(1.5) $u_{1,h} \in K_h$ tel que $p_h u_{1,h} \to \sigma u_1$ dans F fort et $\|u_{1,h}\|_h \leq C$

On suppose que l'approximation est consistante :

$\delta v_{h,k}(t) \in K_h$, $t \in [0,T]$,

$q_h v_{h,k} \rightharpoonup v$ dans $L^\infty(0,T;H)$ faible $*$,

$p_h v_{h,k} \rightharpoonup \sigma v_1$ dans $L^\infty(0,T;F)$ faible $*$,

(1.6)

$p_h v_{h,k}(T) \to \sigma v_2$ et $p_h v_{h,k}(0) = p_h u_{o,h} \to \sigma v_2^o$ dans V faible,

$q_h \delta v_{h,k} \to v_3$ dans $L^\infty(0,T;H)$ faible $*$,

$p_h \delta v_{h,k} \to \sigma v_4$ dans $L^\infty(0,T;F)$ faible $*$,

$q_h \delta v_{h,k}(T-k) \to v_5$ et $q_h \delta v_{h,k}(0) = q_h u_{1,h} \to \sigma v_5^o$ dans H faibl

$q_h \gamma v_{h,k} \to v_6$ dans $L^\infty(0,T;F)$ faible $*$,

on a $v \in \mathcal{K}$ et

$v_1 = \sigma v$, $v_2 = \sigma v(T)$, $v_2^o = \sigma v(0)$, $v_3 = v'$, $v_4 = \sigma v'$, $v_5 = v'(T)$, $v_5^o = v'(0$
$v_6 = v''$.

(mêmes hypothèses dans tout ce qui précède avec δ remplacé par d)

et qu'elle est convergente :

(1.7)

$\forall v \in L^2(0,T;V)$, $v(t) \in K$, $t \in [0,T]$ il existe $v_{h,k}$ avec

$v_{h,k}(t) \in K_h$, $t \in [0,T]$, tel que

$q_h v_{h,k} \to v$ dans $L^2(0,T;H)$ fort,

$p_h v_{h,k} \to \sigma v$ dans $L^2(0,T;F)$ fort.

On supposera (cf. Remarques 7.1, Ch V et 1.3 Ch VI)

(1.8)

K est borné dans V,

K_h est borné dans V_h, indépendamment de h .

1.3 Approximation de $a(u,v)$

Nous supposerons (cf.(7.2)Ch V) que

(1.9) $\quad |A_h u_{o,h}|_h \leq C$, constante indépendante de h

Pour le reste mêmes hypothèses qu'au N°1.3, ch.IV. En particulier que $a(t;u,v)$ et $a_{h,k}(t;u_h,v_h)$ sont symétriques.

1.4 Approximation de $j(t,v)$

On reprend sans changement l'ensemble des hypothèses faites sur la fonctionnelle $(t,v) \to j(t,v)$ au N°1.4; ch.IV. On supposera en outre que (C_1 et C_2 étant deux constantes indépendantes de k,h)

$$(1.10) \qquad \left| \frac{j_h(u_{o,h}) - j_h(v_h)}{k} \right| \leqslant C_1 + C_2 \left| \frac{u_{o,h} - v_h}{k} \right|_{h'} \qquad \forall\, v_h \in K_h \ . \ \blacksquare$$

1.5 Approximation de $f(t)$

On suppose que

$$f, f' \in L^{\infty}(0,T;H).$$

En particulier, on a alors

$$f \in \mathcal{C}\ ([0,T]\ ;\ H).$$

On se donne

$$f_{h,k}(t) = \sum_{i=0}^{N} f_h^i \ \chi_k^i(t)$$

tel que

$$(1.11) \qquad \left| \begin{array}{l} q_h f_{h,k} \to f \ \text{ dans } L^{\infty}(0,T;H) \ \text{ et } \|f_{h,k}\|_h^* \leqslant C \ , \\ \\ q_h\, \delta f_{h,k} \to f' \ \text{ dans } L^{\infty}(0,T;H) \ \text{ et } \|\delta f_{h,k}\|_h^* \leqslant C \ . \ \blacksquare \end{array} \right.$$

1.6 Schémas de discrétisation.

Nous supposerons dans toute la suite de ce chapitre que

$$(A(t)u,v) = a(t;u,v) \ \text{ et } \ j(t,v)$$

ne dépendent pas du temps. On note A_h , $a_h(v_h,v_h), j_h(v_h)$ les éléments discrets correspondants. Pour simplifier, on écrira $a_h(v_h) = a_h(v_h,v_h)$. Nous étudions les deux catégories suivantes de schémas (nous omettons les indices h) :

Schéma I. (Schéma non découplé, 1ere catégorie)
On cherche u^i , $i = 0,1,\ldots,N$ tels que

$$
\left|
\begin{array}{l}
u^0 = u_0 \ , \\
u^1 = u_0 + k u_1 \ , \\
\left(\dfrac{u^{i+1} - 2u^i + u^{i-1}}{k^2} + Au^{i+\theta} - f^i, v - \dfrac{u^{i+1} - u^{i-1}}{2k}\right) + j(v) - \\
- j\left(\dfrac{u^{i+1} - u^{i-1}}{2k}\right) \geqslant 0 \quad , \quad (i \geqslant 1), \\
\forall\ v \in K \quad \text{avec} \quad \dfrac{u^{i+1} - u^{i-1}}{2k} \in K \quad \text{et} \\
u^{i+\theta} = \theta^- u^{i-1} + \theta^0 u^i + \theta^+ u^{i+1} \ , \ \theta^-, \theta^0, \theta^+ \geqslant 0 \ , \ \theta^- + \theta^0 + \theta^+ = 1 \ .
\end{array}
\right.
$$

Schéma II. (Schéma non découplé, 2ème catégorie)
Même inéquation qu'en (1.9) mais en remplaçant

$$
\left|
\begin{array}{l}
\dfrac{u^{i+1} - u^{i-1}}{2k} \quad \text{par} \quad \dfrac{u^{i+1} - u^i}{k} \quad , \\
\text{et} \\
\dfrac{u^{i+1} - u^{i-1}}{2k} \in K \quad \text{par} \quad \dfrac{u^{i+1} - u^i}{k} \in K \ .
\end{array}
\right.
$$

De même que dans le cas parabolique on peut découpler ces schémas.

Schéma III. (Schéma découplé, avec $j=0$)
On cherche u^i , $i=0,1,\ldots,N$ et $u^{i+\frac{1}{2}}$, $i=1,2,\ldots,N$ tels que

$$
\left|
\begin{array}{l}
u^0 = u_0 \ , \\
u^1 = u_0 + k\, u_1 \ , \\
\dfrac{u^{i+\frac{1}{2}} - 2u^i + u^{i-1}}{k^2} + Au^{i+\theta} - f^i = 0 \ , \ (i \geqslant 1) \ ,
\end{array}
\right.
$$

(1) les deux schémas I et II peuvent être regroupés dans une même étude en remplaçant $\dfrac{u^{i+1} - u^{i-1}}{2k}$ dans le schéma I par

$$\frac{u^{i+1} - u^{i-1} - \lambda(u^i - u^{i-1})}{(1-\lambda)} \ , \quad \lambda \in [0,1] \ .$$

$$\frac{u^{i+1} - u^{i-1}}{2k} = P_K\left[\frac{u^{i+\frac{1}{2}} - u^{i-1}}{2k}\right] \quad , \quad (i \geq 1), \ (1)$$

avec

$$u^{i+\theta} = \theta^- u^{i-1} + \theta^o u^i + \theta^* u^{i+\frac{1}{2}} + \theta^+ u^{i+1} \quad ,$$

où l'on se donne

$$\theta^-, \theta^o, \theta^*, \theta^+ \geq 0 \ , \quad \theta^- + \theta^o + \theta^* + \theta^+ = 1 \ .$$

Nous utiliserons les notations

$$\delta^i = \frac{u^{i+1} - u^i}{k} \ ,$$

$$d^i = \frac{u^{i+1} - u^{i-1}}{2k} = \frac{\delta^i + \delta^{i-1}}{2} = \left(\frac{k}{2}\gamma^i + \delta^{i-1}\right),$$

$$\gamma^i = \frac{\delta^i - \delta^{i-1}}{k} = \frac{u^{i+1} - 2u^i + u^{i-1}}{k^2} = 2\left(\frac{d^i - \delta^{i-1}}{k}\right) \quad ,$$

$$u^{i+\theta} = \theta^- u^{i-1} + \theta^o u^i + \theta^+ u^{i+1} \quad ,$$

$$\delta^{i+} = \frac{u^{i+1+\theta} - u^{i+\theta}}{k} = \delta^i + \theta\left[\delta^{i+1} - \delta^i\right] \quad ,$$

$$g^i = \frac{f^{i+1} - f^i}{k} \ .$$

Nous rappelons le Lemme de Gronwald :

Lemme 1.1

Si p_i , i=0,1,...,n, sont des nombres qui vérifient

$$0 \leq p_1 \leq C_1 \quad , \quad \text{constante} \geq 0 \ ,$$

$$0 \leq p_{n+1} \leq C_1 + C_2 \sum_{i=1}^{n} k_i p_i \quad , \quad n = 1, \ldots, N,$$

---

(1) ou bien

$$\frac{u^{i+1} - u^i}{k} = P_K\left[\frac{u^{i+\frac{1}{2}} - u^i}{k}\right] .$$

De façon générale on pourrait définir u^{i+1} par

$$\frac{u^{i+1} - u^{i-1} - \lambda(u^i - u^{i-1})}{(2-\lambda)k} = P_K\left[\frac{u^{i+\frac{1}{2}} - u^{i-1} - \lambda(u^i - u^{i-1})}{(2-\lambda)k}\right] .$$

avec $\lambda \in [0,1]$.

avec C_2 constante ≥ 0, et

$$k_i \geq 0 \ , \quad i=1,\dots,N,$$

$$\sum_{i=1}^{N} k_i = T, \text{ constante positive},$$

alors

$$p_n \leq C_1 \exp(C_2 \sum_{i=1}^{n} k_i) \leq C_1 \exp(C_2 T) \ ,$$

pour $n = 2,3,\dots,N$. ∎

Nous nous servirons en outre du lemme suivant :

Lemme 1.2

Sous les hypothèses (1.8)(1.9)(1.10) on a

$$|\gamma^1|_h^2 \leq C \ ,$$

C étant une constante indépendante de h et k. ∎

Démonstration du Lemme 1.2

Pour le schéma I on remarque

$$(\gamma^1 + A\, u^{1+\theta} - f^1, v-d^1) + j(v) - j(d^1) \geq 0, \qquad \forall\, v \in K,$$

d'où, en prenant $v = \delta^o$, notant que $d^1 - \delta^o = \frac{k}{2}\gamma^1$ et que

$$u^{1+\theta} = u^o + \theta^o k\, \delta^o + \theta^+ k\, \delta^1 = u^o + (\theta^o + \theta^+) k\, \delta^o + \theta^+ k^2 \gamma^1$$

on a

$$(\gamma^1 + A[u^o + (\theta^o + \theta^+) k\, \delta^o + \theta^+ k^2 \gamma^1] - f^1, \tfrac{k}{2}\gamma^1) \leq j(\delta^o) - j(d^1)$$

et, en multipliant par $\frac{2}{k}$ et supprimant le terme $a(\gamma^1, \gamma^1)$ qui est positif, il vient

$$|\gamma^1|^2 \leq |(Au^o, \gamma^1)| + 2(\theta^o + \theta^+)|a(\delta^o, d^1 - \delta^o)| + |(f^1, \gamma^1)| + \frac{2}{k}[j(\delta^o) - j(d^1)] \ .$$

Grâce à (1.8)(1.9)(1.10) on en déduit le résultat. Pour le schéma II, on prend $v = \delta^o$ dans

$$(\gamma^1 + Au^{1+\theta} - f^1, v-\delta^1) + j(v) - j(\delta^1) \geq 0, \quad \forall\, v \in K \ ,$$

et notant que $\delta^1-\delta^0 = k\gamma^1$ et divisant par k on obtient

$$|\gamma^1|^2 \leqslant |(Au^0,\gamma^1)| + (\theta^0+\theta^+)\,|a(\delta^0,\delta^1-\delta^0)| + |(f^1,\gamma^1)| + \tfrac{1}{k}[j(\delta^0) - j(\delta^1)],$$

d'où le résultat grâce aux hypothèses. ∎

1.7 Existence et stabilité pour le schéma I

Nous allons étudier la stabilité et, pour $h,k \to 0$, la convergence du schéma I. Avec les notations du § 1.6, le schéma I s'écrit

$$(1.12)\qquad \left|\begin{array}{l} (\gamma^i, v-d^i) + a(u^{i+\theta}, v-d^i) + j(v) - j(d^i) \geqslant (f^i, v-d^i), (i \geqslant 1), \\ \\ \forall\, v \in K,\ d^i \in K\ . \end{array}\right.$$

L'existence se déduit sans difficulté de la théorie des inéquations elliptiques en mettant (1.12) sous la forme

$$((I+\theta^+ k^2 A)d^i - \tilde{f}^i,\ v-d^i) + k\,[j(v) - j(d^i)] \geqslant 0\ ,\quad \forall\, v \in K,\ d^i \in K,$$

où $\tilde{f}^i$ rassemble les termes connus.

Nous allons chercher des majorations convenables de

$$u_{h,k} = \sum_{i=0}^{N} u_h^i\ \chi_k^i$$

et de ses dérivées discrètes.

(Rappelons que nous omettons les indices h autant que possible). Pour cela nous écrivons (1.12) avec i remplacé par $i+1$, soit

$$(1.13)\qquad \left|\begin{array}{l} (\gamma^{i+1}, v-d^{i+1}) + a(u^{i+1+\theta}, v-d^{i+1}) + j(v) - j(d^{i+1}) \geqslant (f^{i+1}, v-d^{i+1}), (i \geqslant \\ \\ \forall\, v \in K,\ d^{i+1} \in K\ . \end{array}\right.$$

Prenant $v=d^{i+1}$ dans (1.12) et $v=d^i$ dans (1.13) et additionnant il vient

$$(1.14)\qquad (\gamma^{i+1} - \gamma^i, d^{i+1}-d^i) + a(u^{i+1+\theta} - u^{i+\theta}, d^{i+1}-d^i) \leqslant (f^{i+1}-f^i, d^{i+1}-d^i).$$

Or

$$d^{i+1} - d^i = \frac{\delta^{i+1} - \delta^{i-1}}{2} = \frac{k}{2}\ (\gamma^{i+1} + \gamma^i),$$

d'où, après multiplication par $\frac{2}{k}$,

(1.15) $$(\gamma^{i+1} - \gamma^i,\ \gamma^{i+1} + \gamma^i) + a(\delta^{i+\theta},\ \delta^{i+1} - \delta^{i-1}) \leqslant k(g^i,\ \gamma^{i+1} + \gamma^i),$$

mais en notant que $a(,)$ est symétrique, on a

(1.16) $$a(\delta^{i+1},\ \delta^{i+1} - \delta^{i-1}) = \tfrac{1}{2}\left[a(\delta^{i+1}) - a(\delta^{i-1}) + a(\delta^{i+1} - \delta^{i-1})\right],$$

(1.17) $$a(\delta^{i},\ \delta^{i+1} - \delta^{i-1}) = \tfrac{1}{2}\left[a(\delta^{i+1}) - a(\delta^{i-1}) + a(\delta^{i} - \delta^{i-1}) - a(\delta^{i+1} - \delta\right.$$

(1.18) $$a(\delta^{i-1},\ \delta^{i+1} - \delta^{i-1}) = \tfrac{1}{2}\left[a(\delta^{i+1}) - a(\delta^{i-1}) - a(\delta^{i+1} - \delta^{i-1})\right] \quad ,$$

et en reportant dans (1.15) il vient

(1.19) $$|\gamma^{i+1}|^2 - |\gamma^i|^2 + \tfrac{1}{2}\,a(\delta^{i+1}) - \tfrac{1}{2}\,a(\delta^{i-1}) + \frac{\theta^o}{2}\left[a(\delta^i - \delta^{i-1}) - a(\delta^{i+1} - \delta^i)\right] + \frac{\theta^+ - \theta^-}{2}\,a(\delta^{i+1} - \delta^{i-1}) \leqslant k(g^i,\ \gamma^{i+1} + \gamma^i)\,.$$

En sommant de $i=1$ à $n \leqslant N-2$ on a

(1.20) $$|\gamma^{n+1}|^2 - |\gamma^1|^2 + \tfrac{1}{2}\,a(\delta^{n+1}) + \tfrac{1}{2}\,a(\delta^n) - \tfrac{1}{2}\,a(\delta^1) - \tfrac{1}{2}\,a(\delta^o) + \frac{\theta^o}{2}\left[a(\delta^1 - \delta^o) - a(\delta^{n+1} - \delta^n)\right] + \frac{\theta^+ - \theta^-}{2}\sum_{i=1}^{n} a(\delta^{i+1} - \delta^{i-1}) \leqslant$$

$$\leqslant \sum_{i=1}^{n} C\,k\left(|\gamma^{i+1}| + |\gamma^i|\right) \leqslant C + \sum_{i=1}^{n} k\,|\gamma^i|^2 + \tfrac{1}{2}\,|\gamma^{n+1}|^2\,.$$

Or

(1.21) $$\frac{\theta^o}{2}\,a(\delta^1 - \delta^o) \leqslant a(\delta^1) + a(\delta^o) \quad , \quad (\max\ 0^o = 1)\,,$$

(1.22) $$\frac{\theta^o}{2}\,a(\delta^{n+1} - \delta^n) \leqslant \frac{\theta^o}{2}\,C\,k^2 S(h)^2\,|\gamma^{n+1}|^2$$

(1.23) $$\tfrac{1}{2}\sum_{i=1}^{n} a(\delta^{i+1} - \delta^{i-1}) = \sum_{i=1}^{n}\left[a(\delta^{i+1} - \delta^i) + a(\delta^i - \delta^{i-1})\right] \leqslant$$

$$\leqslant 2\sum_{i=1}^{n} a(\delta^i - \delta^{i-1}) + a(\delta^{n+1} - \delta^n) \leqslant$$

$$\leqslant 2\sum_{i=1}^{n} Ck^2 S(h)^2\,|\gamma^i|^2 + Ck^2 S(h)^2\,|\gamma^{n+1}|^2\,,$$

d'où, avec $(\theta^- - \theta^+)^+ = \max\ [0,\ \theta^- - \theta^+]$

$$(1.24)\quad \left[\tfrac{1}{2} - \left(\frac{\theta^o}{2} - (\theta^- - \theta^+)^+\right)Ck^2S(h)^2\right] |\gamma^{n+1}|^2 + \tfrac{1}{2}\, a(\delta^{n+1}) + \tfrac{1}{2}\, a(\delta^n) \leqslant C +$$

$$+ \sum_{i=1}^{n} \left[1 + (\theta^- - \theta^+)^+ 2CkS(h)^2\right] k\, |\gamma^i|^2 + 2\, a(\delta^1) + 2\, a(\delta^o) + |\gamma^1|^2 .$$

Avec les hypothèses (1.8) $(a(\delta^1) \leqslant C)$, (1.5) $(a(\delta^o) \leqslant C)$, et le Lemme 1.2 $(|\gamma^1| \leqslant C)$ on déduit, avec le Lemme de Gronwald le théorème suivant

Théorème 1.1

Pour le schéma I, lorsque $(h,k) \to 0$

(1.25) - de façon quelconque si $\theta^o = 0, \quad \theta^+ \geqslant \theta^-$,

- avec les restrictions (dites conditions de stabilité)

$$(1.26)\quad \begin{cases} 1 - Ck^2S(h)^2 \geqslant \beta > 0, \quad \beta \text{ fixe} \quad \in\ ]0,1[\ , \\ \text{si} \quad \theta^o > 0 \ \text{ et } \ \theta^+ \geqslant \theta^- \end{cases}$$

$$(1.27)\quad \begin{cases} 1 - Ck^2S(h)^2 \geqslant \beta > 0, \quad \beta \text{ fixe} \quad \in\ ]0,1[\\ kS(h)^2 \leqslant C, \ \text{constante fixe quelconque,} \\ \text{si} \quad \theta^o \geqslant 0 \ \text{ et } \ \theta^- > \theta^* \ , \end{cases}$$

on a, pour tout $n=0,1,\ldots$

$$(1.28)\quad \|u_h^n\|_h \leqslant C$$

$$(1.29)\quad \left\| \frac{u_h^{n+1} - u_h^{n-1}}{2k} \right\|_h \leqslant C \ , \qquad \left\| \frac{u_h^{n+1} - u_h^n}{k} \right\|_h \leqslant C \ ,$$

$$(1.30)\quad \left| \frac{u_h^{n+1} - 2u_h^n + u_h^{n-1}}{k^2} \right|_h \leqslant C \ . \ \blacksquare$$

Le résultat (1.28) se déduit de (1.29) grâce à (1.5) et $u^{n+1} = u^n + k\ \delta^n$. $\blacksquare$

1.8 Convergence faible du schéma I.

D'après les hypothèses (1.3)(1.4)(1.5)(1.6) et dans les conditions du Théorème 1.1 la suite

$$u_{h,k} = \sum u_h^i \quad \chi_k^i$$

vérifie, pour une sous-suite extraite notée encore $u_{h,k}$,

(1.31)
$$\left|\begin{array}{l} q_h u_{h,k} \to \bar{u} \text{ dans } L^\infty(0,T;H) \text{ faible } * \text{ , } \bar{u} \in \mathcal{K} \text{ ,} \\ p_h u_{h,k} \to \sigma\bar{u} \text{ dans } L^\infty(0,T;F) \text{ faible*,} \\ p_h du_{h,k} \to \sigma\bar{u}' \text{ dans } L^\infty(0,T;F) \text{ faible*,} \\ q_h \gamma u_{h,k} \to u'' \text{ dans } L^\infty(0,T;H) \text{ faible*,} \\ p_h u_{h,k}(T) \to \sigma\bar{u}(T) \text{ et } p_h u_{h,k}(o) \to \sigma\bar{u}(0) \text{ dans V faible,} \\ q_h \delta u_{h,k}(T-k) \to u'(T) \text{ et } q_h \delta u_{h,k}(0) \to \sigma\bar{u}'(0) \text{ dans H faible.} \end{array}\right.$$

Soit alors v quelconque, $v \in L^2(0,T;V)$, $v(t) \in K$, p.p. $t \in [0,T]$ et une suite $v_{h,k}$ qui converge vers v au sens (1.7). En remplaçant v par v_h^i dans (9.14) on obtient

(1.32)
$$\left(\frac{u_h^{i+1} - 2u_h^i + u_h^{i-1}}{k^2} + A_h u_h^{i+\theta} \text{ , } v_h^i\right)_h - \left(f_h^i, v_h - \frac{u_h^{i+1} - u_h^{i-1}}{2k}\right) + j_h(v_h^i) \geq D_{h,k}^i =$$
$$= \left(\frac{u_h^{i+1} - 2u_h^i + u_h^{i-1}}{k^2} + A_h u_h^{i+\theta} \text{ , } \frac{u_h^{i+1} - u_h^{i-1}}{2k}\right)_h + j_h\left(\frac{u_h^{i+1} - u_h^{i-1}}{2k}\right) .$$

Nous allons montrer qu'en multipliant par k et en sommant de i=1 à N-1 on obtient

(1.33)
$$\int_0^T (\bar{u}'' + A\bar{u}, v)dt - \int_0^T (f, v-\bar{u}')dt + \int_0^T j(v)dt \geq \int_0^T (\bar{u}''+A\bar{u}, \bar{u}')dt + \int_0^T j(\bar{u}')dt.$$

ce qui établira que $\bar{u}$ est solution du problème. L'obtention du membre de gauche de (1.33) étant immédiate il reste à examiner le deuxième membre de (1.32).

En multipliant par $2k$ le membre de droite de (1.32) on obtient

$$(1.34)\qquad 2kD^i_{h,k} = (\delta^i - \delta^{i-1}\ ,\ \delta^i + \delta^{i-1})_h + a_h(u_h^{i+\theta}\ ,\ u_h^{i+1} - u_h^{i-1}) + j_h(d^i) =$$

$$= |\delta^i|^2_h - |\delta^{i-1}|^2_h + a_h(u_h^{i+\theta}\ ,\ u_h^{i+1} - u_h^{i-1}) + 2k\, j_h(d^i)\ .$$

Mais (omettons les indices h) grâce à (1.16)(1.17)(1.18) on obtient, en sommant de $i=1$ à $N-1$,

$$(1.35)\qquad \begin{aligned} 2D_{h,k} = \sum_{i=1}^{N-1} 2k\, D^i_{h,k} = {} & |\delta^{N-1}|^2 - |\delta^0|^2 + \tfrac{1}{2}a(u^N) + \tfrac{1}{2}a(u^{N-1}) - \tfrac{1}{2}a(u^1) - \tfrac{1}{2}a(u^0) \\ & + \frac{\theta^0}{2}\left[a(u^1 - u^0) - a(u^N - u^{N-1})\right] + \frac{\theta^+ - \theta^-}{2}\sum_{i=1}^{N-1} a(u^{i+1} - u^{i-1}) + 2\sum_{i=1}^{N-1} j_h(d^i)k \end{aligned}$$

Or on a les résultats suivants

$$(1.36)\qquad \liminf |\delta^{N-1}|^2 \geqslant |u'(T)|^2 \qquad \text{(d'après (1.31)),}$$

$$(1.37)\qquad |\delta^0|^2 \quad \to \quad |u_1|^2 = |u'(0)|^2 \ \text{(d'après (1.4) et (1.31)),}$$

$$\left.\begin{aligned} (1.38)\qquad & \liminf a(u^N) \geqslant a(\bar{u}(T)) \\ (1.39)\qquad & \liminf a(u^{N-1}) \geqslant a(\bar{u}(T)) \end{aligned}\right\} \text{(d'après la consistance de a(,) et (1.31}$$

$$\left.\begin{aligned} (1.40)\qquad & a(u^1) \quad \to \quad a(u_o) = a(\bar{u}(0)) \\ (1.41)\qquad & a(u^0) \quad \to \quad a(u_o) = a(\bar{u}(0)) \end{aligned}\right\} \text{(d'après la convergence de a(,) et (1.31)}$$

$$(1.42)\qquad \liminf 2\sum_{i=1}^{N-1} j_h(d^i)k \geqslant 2\int_0^T j(\bar{u}')dt \qquad \text{(d'après la consistance de j}$$

d'où

$$(1.43)\qquad \liminf 2\, D_{h,k} \geqslant 2D = |\bar{u}'(T)|^2 - |\bar{u}'(0)|^2 + a(\bar{u}(T)) - a(\bar{u}(0)) + 2\int_0^T j(\bar{u}')dt +$$

$$+ \liminf r\ ,$$

avec

$$(1.44)\qquad r = \tfrac{1}{2}\left\{ \theta^{o}\left[a(u^{1}-u^{o}) - a(u^{N}-u^{N-1})\right] + (\theta^{+}-\theta^{-})\sum_{i=1}^{N-1} a(u^{i+1}-u^{i-1})\right\} .$$

D'après la formule de Green (Lemme 2.4 ch. III et Lemme 2.6 ch. V) on a en divisant par 2

$$(1.45)\qquad D = \int_0^T (\bar{u}'',\bar{u}')dt + \int_0^T a(\bar{u},\bar{u}')dt + \int_0^T j(\bar{u}')dt + \liminf \frac{r}{2} \quad .$$

En sommant et en passant à la limite dans (1.32) on obtient alors

$$(1.46)\qquad \int_0^T (\bar{u}'' + A\bar{u} - f,\ v - \bar{u}')dt + \int_0^T [j(v) - j(\bar{u}')]dt \geqslant \liminf \frac{r}{2} \quad .$$

Remarquons dès l'abord qu'avec $\theta^{o} = 0$ et $\theta^{+} > \theta^{-}$ on a $\liminf \frac{r}{2} \geqslant 0$ et $\bar{u}$ est bien solution de l'inéquation initiale (1.1) . Dans les autres cas on a

$$(1.47)\qquad r = \tfrac{1}{2}\left\{ \theta^{o}k^{2}\left[a(\delta^{o}) - a(\delta^{N-1})\right] + (\theta^{+}-\theta^{-})\sum_{i=1}^{N-1} 4k^{2}a(d^{i}) \right. .$$

Or

$$(1.48)\qquad a(\delta^{o}) \leqslant C \ \|\delta^{o}\|_{h}^{2} \leqslant C \quad , \quad \text{(par hypothèse (1.4))} \ ,$$

$$(1.49)\qquad a(\delta^{N-1}) \leqslant C \ \|\delta^{N-1}\|_{h}^{2} \leqslant C \quad , \ \text{(d'après (1.29))},$$

$$(1.50)\qquad \sum_{i=1}^{N-1} k^{2}a(d^{i}) \leqslant k \sum_{i=1}^{N-1} k\ C \leqslant k\ C\ T \quad , \ \text{(d'après (1.29))},$$

et l'on a $\liminf |r| = 0$ d'où, évidemment,

$$(1.51)\qquad \liminf r = 0 \ ,$$

ce qui démontre (voir (1.46) avec (1.51)) que $\bar{u}$ est bien solution de l'inéquation initiale. La solution $\bar{u}$ étant unique, il n'est pas nécessaire d'extraire une sous suite, et la suite toute entière converge vers u . On peut énoncer le

Théorème 1.2

Pour le schéma I, lorsque $(h,k) \to 0$

(1.52) - de façon quelconque si $\theta^o = 0$, $\theta^+ \geqslant \theta^-$

- avec les restrictions (dites conditions de stabilité)

(1.53)
$$\begin{cases} 1 - Ck^2 S(h)^2 \geqslant \beta > 0 \ , \quad \beta \text{ fixe} \quad \in]0,1[\ , \\ \text{si} \quad \theta^o > 0 \ \text{ et } \quad \theta^- \leqslant \theta^+ \ ; \end{cases}$$

(1.54)
$$\begin{cases} 1 - Ck^2 S(h)^2 \geqslant \beta > 0 \ , \quad \beta \text{ fixe} \quad \in]0,1[\ , \\ k\, S(h)^2 \leqslant C \ , \quad \text{constante fixe quelconque,} \\ \text{si} \quad \theta^o \geqslant 0 \quad \text{et} \quad \theta^- > \theta^+ \ , \end{cases}$$

on a, avec u solution de l'inéquation (1.1),

(1.55)
$$\begin{cases} q_h u_{h,k} \to u \quad \text{dans} \quad L^\infty(0,T;H) \text{ faible*}, \\ p_h u_{h,k} \to \sigma u \quad \text{dans} \quad L^\infty(0,T;F) \text{ faible*}, \\ p_h d u_{h,k} \to \sigma u' \quad \text{dans} \quad L^\infty(0,T;F) \text{ faible*}, \text{(1)} \\ q_h \gamma u_{h,k} \to u'' \quad \text{dans} \quad L^\infty(0,T;H) \text{ faible*}. \end{cases}$$

Remarque 1.1

La démonstration précédente prouve l'existence de $u \in \mathcal{K}$ solution de (1.1). ∎

1.9 Existence et stabilité pour le schéma II

Avec les notations du §1.6, le schéma II s'écrit

(1.56)
$$\begin{cases} (\gamma^i, v - \delta^i) + a(u^{i+\theta}, v - \delta^i) + j(v) - j(\delta^i) \geqslant (f^i, v - \delta^i), \ (i \geqslant 1), \\ \forall\, v \in K \ , \quad \delta^i \in K \ . \end{cases}$$

(1) ou bien d remplacé par δ .

L'existence se déduit de la théorie des inéquations elliptiques en mettant (1.12) sous la forme

$$((I + \theta^+ k^2 A)\,\delta^i - \tilde{f}^i\ ,\ v - \delta^i) + k\,[j(v) - j(\delta^i)] \geqslant 0\ ,\quad \forall v \in K,\quad \delta^i \in K\ ,$$

où $\tilde{f}^i$ rassemble les termes connus.

Cherchant des majorations convenables de $u_{h,k}$ et de ses dérivées discrètes, nous écrivons (1.56) avec i remplacé par $i+1$ soit

$$(1.57)\qquad \left|\begin{array}{l} (\gamma^{i+1}\ ,\ v-\delta^{i+1}) + a(u^{i+1+\theta}\ ,\ v-\delta^{i+1}) + j(v) - j(\delta^{i+1}) \geqslant (f^{i+1}, v-\delta^{i+1}), \\ \\ \forall v \in K,\quad \delta^{i+1} \in K\ . \end{array}\right.$$

Prenant $v = \delta^{i+1}$ dans (1.56) et $v = \delta^i$ dans (1.57), additionnant et divisant par k, il vient

$$(1.58)\qquad (\gamma^{i+1} - \gamma^i,\ \gamma^{i+1}) + a(\delta^{i+\theta}\ ,\ \delta^{i+1} - \delta^i) \leqslant (\frac{f^{i+1} - f^i}{k},\ \delta^{i+1} - \delta^i)\ .$$

Or

$$(1.59)\qquad a(\delta^{i+1}\ ,\ \delta^{i+1} - \delta^i) = \tfrac{1}{2}\,[a(\delta^{i+1}) - a(\delta^i) + a(\delta^{i+1} - \delta^i)]\ ,$$

$$(1.60)\qquad a(\delta^i\ ,\ \delta^{i+1} - \delta^i) = \tfrac{1}{2}\,[a(\delta^{i+1}) - a(\delta^i) - a(\delta^{i+1} - \delta^i)]\ ,$$

$$(1.61)\qquad a(\delta^{i-1}\ ,\ \delta^{i+1} - \delta^i) = \tfrac{1}{2}\,[a(\delta^{i+1}) - a(\delta^i) - a(\delta^{i+1} - \delta^i) - 2a(\delta^{i+1} - \delta^i, \delta^i - \delta^{i-1}$$

et en reportant dans (1.58) il vient, après multiplication par 2,

$$(1.62)\qquad |\gamma^{i+1}|^2 - |\gamma^i|^2 + |\gamma^{i+1} - \gamma^i|^2 + a(\delta^{i+1}) - a(\delta^i) + (\theta^+ - \theta^\circ - \theta^-)a(\delta^{i+1} - \delta^i)$$
$$\leqslant k(g^i, \gamma^{i+1}) + 2\,\theta^- a(\delta^{i+1} - \delta^i,\ \delta^i - \delta^{i-1}) \leqslant k(g^i,\ \gamma^{i+1}) + \theta^- a(\delta^{i+1} - \delta^i) +$$
$$+\ \theta^- a(\delta^i - \delta^{i-1})\ ,$$

$$(1.63)\qquad |\gamma^{i+1}|^2 - |\gamma^i|^2 + |\gamma^{i+1} - \gamma^i|^2 + a(\delta^{i+1}) - a(\delta^i) + (\theta^+ - \theta^\circ - 2\,\theta^-)$$
$$a(\delta^{i+1} - \delta^i) \leqslant k(g^i,\ \gamma^{i+1}) + \theta^-\, a(\delta^i - \delta^{i-1})$$

et, par sommation de $i=1$ à $n \leqslant N-1$, en supprimant le terme en $|\gamma^{i+1} - \gamma^i|$ et en notant que

(1.64) $$\sum_{i=1}^{n} a(\delta^i - \delta^{i-1}) = \sum_{i=0}^{n} a(\delta^{i+1} - \delta^i)$$

on obtient

(1.65) $$|\gamma^{n+1}|^2 - |\gamma^1|^2 + a(\delta^{n+1}) - a(\delta^1) + (\theta^+ - \theta^\circ - 3\theta^-) \sum_{i=1}^{n} a(\delta^{i+1} - \delta^i) \leqslant$$

$$\leqslant \sum_{i=1}^{n} Ck\,|\gamma^{i+1}| + \theta^- a(\delta^1 - \delta^\circ) .$$

Or

$$|\gamma^1| \leqslant C \quad \text{d'après le Lemme 1.2,}$$

$$a(\delta^1) \leqslant C \quad \text{d'après (1.8),}$$

$$a(\delta^1 - \delta^\circ) \leqslant C \quad \text{d'après (1.8)(1.5),}$$

et

$$\sum_{i=1}^{n} Ck\,|\gamma^{i+1}| \leqslant \sum_{i=1}^{n-1} \frac{kC}{2}(1 + |\gamma^{i+1}|^2) + \frac{k^2C^2}{2} + \frac{|\gamma^{n+1}|^2}{2} ,$$

$$\sum_{i=1}^{n} a(\delta^{i+1} - \delta^i) \leqslant \sum_{i=1}^{n} Ck^2 S(h)^2\,|\gamma^{i+1}|^2 .$$

D'où

(1.66) $$\left(\tfrac{1}{2} - \max[0,\ \theta^\circ + 3\theta^- - \theta^+]\,Ck^2 S(h)^2\right)\,|\gamma^{n+1}|^2 + a(\delta^{n+1}) \leqslant$$

$$\leqslant \sum_{i=1}^{n-1} kC\left(\tfrac{1}{2} + \max[0,\ \theta^\circ + 3\theta^- - \theta^+]\,kS(h)^2\right)\,|\gamma^{i+1}|^2 + C .$$

Par application du Lemme de Gronwald, on obtient alors le

Théorème 1.3

Pour le schéma II, lorsque $(h,k) \to 0$

- de façon quelconque si $\theta^+ - \theta^\circ - 3\theta^- \geqslant 0$

- dans les autres cas avec les conditions

(1.67)
$$\begin{cases} 1 - Ck^2 S(h)^2 \geqslant \beta > 0 , & \beta \text{ fixe } \in \ ]0,1[, \\ kS(h)^2 \leqslant C , & \text{constante quelconque indépendante de h et k,} \end{cases}$$

on a, pour tout $n=0,1,\ldots$ les majorations (1.28)(1.29)(1.30). ∎

Remarque 1.2

Dans le cas $\theta^+ - \theta^0 - 3\theta^- \geqslant 0$ la stabilité est inconditionnelle. La stabilité est donc inconditionnelle dans les cas suivants

$$\theta^0 = 0 , \ \theta^+ \geqslant 3\theta^- , \ (\text{i.e. } \theta^+ \geqslant \tfrac{3}{4} , \ \theta^- \leqslant \tfrac{1}{4})$$

$$\theta^- = 0 \qquad \theta^+ \geqslant \theta^0 \quad (\text{i.e. } \theta^+ \geqslant \tfrac{1}{2} , \ \theta^0 \leqslant \tfrac{1}{2})$$

La stabilité est conditionnelle si $\theta^+ - \theta^0 - 3\theta^- < 0$, par exemple dans tous les cas où $\theta^+ = 0$. ∎

1.10 Convergence faible du schéma II.

Avec les hypothèses faites, il existe alors une sous-suite $u_{h,k}$ et un élément $\bar{u} \in \mathcal{K}$ tel que l'on ait les convergences (1.31). En prenant v et $v_{h,k}$ comme au N° 1.8 on a (nous omettons les indices h)

$$\left(\frac{u^{i+1} - 2u^i + u^{i-1}}{k^2} + Au^{i+\theta} , v^i\right) - \left(f^i, v - \frac{u^{i+1} - u^i}{k}\right) + j(v^i) \geqslant D^i_{h,k} =$$

$$= \left(\frac{u^{i+1} - 2u^i + u^{i-1}}{k^2} + Au^{i+\theta} , \frac{u^{i+1} - u^i}{k}\right) + j\left(\frac{u^{i+1} - u^i}{k}\right) .$$

Nous allons voir qu'en multipliant par k et en sommant de $i=1$ à $N-1$ on obtient (1.33). Examinons $D^i_{h,k}$. En multipliant par $2k$, il vient

$$2k\, D^i_{h,k} = 2(\delta^i - \delta^{i-1} , \delta^i) + 2a(u^{i+\theta} , u^{i+1} - u^i) + 2kj(\delta^i).$$

Utilisant (1.59)(1.60)(1.61), on obtient par sommation de $i=1$ à $N-1$,

$$2\, D_{h,k} = \sum_{i=1}^{N-1} 2kD^i_{h,k} = |\delta^{N-1}|^2 - |\delta^0|^2 + a(u^N) - a(u^1) + \sum_{i=1}^{N-1} (\theta^+ - \theta^0 - \theta^-)a(u^{i+1} -$$

$$\ldots 2\theta^- \sum_{i=1}^{N-1} a(u^{i+1} - u^i , u^i - u^{i-1}) + 2\sum_{i=1}^{N-1} j_k(\delta^i) .$$

Avec (1.36)(1.37)(1.38)(1.40)(1.42) et comme les termes en $a(u^{i+1} - u^i)$ et $a(u^{i+1} - u^i, u^i - u^{i-1})$ tendent vers 0 en $0(k^2)$ (d'après (1.29)) on obtient

$$\liminf D_{h,k} \geqslant 2D = |u'(T)|^2 - |u'(0)|^2 + a(\bar{u}(T)) - a(\bar{u}(0)) + 2\int_0^T j(\bar{u}')dt .$$

On obtient alors sans difficulté (raisonnement semblable à celui de (1.8)) le

Théorème 1.4

Pour le schéma II, lorsque $(h,k) \to 0$, de façon quelconque si $\theta^+ - \theta^0 - 3\theta^- \geqslant 0$ et avec (1.67) dans les autres cas on a les convergences faibles (1.55).(1) ∎

1.11 Convergence forte des schémas I et II.

Nous allons établir dans ce dernier paragraphe les deux théorèmes :

Théorème 1.5

Pour le schéma I, lorsque $(h,k) \to 0$

- de façon quelconque si $\theta^0 = 0$, $\theta^+ \geqslant \theta^-$,
- avec les restrictions (1.53) ou (1.54) dans les autres cas,

on a,

$$(1.68)\quad \begin{cases} p_h u_{h,k}(t) \to \sigma u(t) \text{ dans } V \text{ fort}, & t \in [0,T] , \\ q_h \delta u_{h,k}(t) \to u'(t) \text{ dans } H \text{ fort}, & t \in [0,T] . \end{cases}$$ ∎

Théorème 1.6

Pour le schéma II, lorsque $(h,k) \to 0$,

- de façon quelconque si $\theta^+ - \theta^0 - 3\theta^- \geqslant 0$,
- avec (1.67) dans les autres cas,

on a les convergences fortes (1.68). ∎

---

(1) Remplaçer dans (1.55) $p_h du_{h,k}$ par $p_h \delta u_{h,k}$. ∎

Démonstration du Théorème 1.5 (Schéma I)

Nous allons évaluer la quantité

$$X(t) = |u' - q_h \delta u_{h,k}|^2 + \tilde{a}(u(t) - p_h u_{h,k}(t)), \tag{1.69}$$

que nous écrivons

$$X(t) = X_1(t) - 2X_2(t) + X_3(t), \tag{1.70}$$

avec

$$\left| \begin{array}{l} X_1(t) = |u'|^2 + \tilde{a}(u), \\ X_2(t) = (u', \delta u_{h,k}) + \tilde{a}(u, p_h u_{h,k}), \\ X_3(t) = |\delta u_{h,k}|^2 + \tilde{a}(p_h u_{h,k}). \end{array} \right. \tag{1.71}$$

1er terme : $X_1(t)$

Avec les hypothèses et les Lemmes d'intégration par partie on a

$$X_1(t) = 2 \int_0^t (u'',u')ds + 2 \int_0^t a(u,u')ds + a(u_o) + |u'(0)|^2 . \tag{1.72}$$

2ème terme : $X_2(t)$

Soit nk tel que $t \in [nk,(n+1)k[$ et posons

$$\tau_k u = \sum_{i=0}^{N-1} u^{i+1} \chi_k^i \quad , \text{ avec } u^i = u(ik). \tag{1.73}$$

Rappelons que $u_{h,k} = \sum_{i=0}^{N} u_h^i \chi_k^i$ et que $a(,)$ est symétrique et ne dépend pas du temps. On obtient aisément (on omet p_h et on note $u^i = u(ik)$)

$$\tilde{a}(u(t),p_h u_{h,k}(t)) = \int_{nk}^t \tilde{a}(u',u_h^n) + \tilde{a}(u^n,u_h^n) .$$

Mais pour $i=1,\ldots,N$, on a

$$\int_{(i-1)k}^{ik} \tilde{a}(u',u_h^{i-1})ds = \tilde{a}(u^i,u_h^{i-1}) - \tilde{a}(u^{i-1},u_h^n),$$

d'où, en ajoutant $a(u^i,u_h^i)$ de deux côtés,

$$\tilde{a}(u^i,u_h^i) = \int_{(i-1)k}^{ik} \tilde{a}(u',u_h^{i-1})ds + \tilde{a}(u^i,u_h^i - u_h^{i-1}) + \tilde{a}(u^{i-1},u_h^{i-1}),$$

et finalement

$$(1.74)\qquad \tilde{a}(u(t),p_h u_{h,k}(t)) = \int_0^t \tilde{a}(u',p_h u_{h,k})ds + \int_0^{nk} \tilde{a}(\tau_k u,p_h \delta u_{h,k})ds +$$

$$+ \tilde{a}(u(0),u_{h,k}(0)).$$

Toujours pour le second terme, évaluons maintenant $(u',\delta u_{h,k})$. On a

$$(u',\delta u_{h,k}) = \int_{nk}^t (u'',\delta^n)ds + (u'(nk),\delta^n)\ .$$

Mais

$$\int_{(i-1)k}^{ik} (u'',\delta^{i-1}) = (u'(ik),\delta^{i-1}) - (u'((i-1)k),\delta^{i-1}),$$

et, en ajoutant $(u'(ik),\delta^{i-1})$ des 2 côtés,

$$(u'(ik),\delta^i) = \int_{(i-1)k}^{ik} (u'',\delta^{i-1})ds + \int_{(i-1)k}^{ik} (u'(ik),\gamma^i)ds +$$

$$+ (u'((i-1)k),\delta^{i-1}),$$

d'où

$$(1.75)\quad (u',\delta u_{h,k}) = \int_0^t (u'',q_h \delta u_{h,k})ds + \int_0^{nk} (\tau_k u',q_h \gamma u_{h,k})ds + (u'(0),q_h \delta^o)\ .$$

On peut alors passer à la limite dans X_2 et l'on obtient

$$(1.76)\qquad \lim\ X_2(t) = X_1(t). \qquad \blacksquare$$

3ème terme : $X_3(t)$

On utilise les notations

$$\left|\begin{array}{l}
u_{h,k}(t) = \sum_{i=0}^{N} u_h^i \ \chi_k^i(t), \\[2ex]
u_{h,k}^{\theta}(t) = \sum_{i=0}^{N-1} [\theta^- u_h^{i-1} + \theta^0 u_h^i + \theta^+ u_h^{i+1}] \ \chi_k^i(t), \\[2ex]
\delta u_{h,k}(t) = \sum_{i=1}^{N-1} \left(\frac{u_h^{i+1} - u_h^i}{k}\right) \chi_k^i(t), \\[2ex]
du_{h,k}(t) = \sum_{i=1}^{N-1} \left(\frac{u_h^{i+1} - u_h^{i-1}}{2k}\right) \chi_k^i(t), \\[2ex]
\gamma u_{h,k}(t) = \sum_{i=1}^{N-1} \left(\frac{u_h^{i+1} - 2u_h^i + u_h^{i-1}}{k^2}\right) \chi_k^i(t) .
\end{array}\right.$$

En écrivant (1.12) sous forme intégrée, pour $t \in [k,T-k]$, on a

$$(1.77) \quad \int_k^t (\gamma v_{h,k} + A_h u_{h,k}^{\theta}, du_{h,k}) ds \leqslant \int_k^t (\gamma u_{h,k} + A_h u_{h,k}^{\theta}, v_{h,k}) ds +$$

$$+ \int_k^t (f_{h,k}, du_{h,k} - v_{h,k}) ds + \int_k^t [j_h(v_{h,k}) - j_h(du_{h,k})] \, ds$$

où l'on prend $v_{h,k} = \sum_{i=1}^{N-1} v_h^i \ \chi_k^i$ tel que

$$(1.78) \quad p_h v_{h,k} \to \sigma u' \text{ dans } L^2(0,T;F) \text{ fort et } v_h^i \in K_h .$$

Mais pour $t \in [nk,(n+1)k[$, $n \geqslant 1$, on a

$$(1.79) \quad \int_k^t = \int_k^{(n+1)k} - \int_t^{(n+1)k} = \sum_{i=1}^{n} \int_{ik}^{(i+1)k} - \int_t^{(n+1)k} .$$

(On remarquera que le terme sous la dernière intégrale tend vers 0 avec k).

D'autre part

$$(1.80)\begin{cases} \displaystyle\int_{ik}^{(i+1)k} (\gamma u_{h,k} + A_h u^{\theta}_{h,k}\ ,\ du_{h,k})ds = (\frac{\delta^i - \delta^{i-1}}{k} + Au^{i+\theta}\ ,\ \frac{\delta^i + \delta^{i-1}}{2})k = \\ \\ = \frac{1}{2}\ \{\ |\delta^i|^2 - |\delta^{i-1}|^2 + (Au^{i+\theta}\ ,u^{i+1} - u^{i-1})\ \}\ , \end{cases}$$

d'où, en utilisant (1.78)(1.79) et (1.16)(1.17)(1.18), et avec $\varepsilon(k) \to 0$ avec k,

$$(1.81)\begin{cases} 2 \displaystyle\int_k^t (\gamma u_{h,k} + A_h u^{\theta}_{h,k}, du_{h,k})ds = \varepsilon(k) + |\delta u_{h,k}(t)|^2 - |\delta u_{h,k}(0)|^2 + \\ \\ + \frac{1}{2}\{\ a(u_{h,k}(t+k)) + a(u_{h,k}(t)) - a(u_{h,k}(k)) - a(u_{h,k}(0)) + \theta^o[a(u^1 - u^o) - \\ \\ -a(u^{n+1} - u^n)\ ] + (\theta^+ - \theta^-) \sum_{i=1}^{n} a(u^{i+1} - u^{i-1})\} \quad . \end{cases}$$

On a alors avec $\frac{1}{2}\, a(u_{h,k}(t)) = (1 - \frac{1}{2})a(u_{h,k}(t))$, et reportant (1.81) dans (1.77), il vient avec $\varepsilon'(k) \to 0$, $k \to 0$,

$$(1.82)\begin{cases} |\delta u_{h,k}(t)|^2 + a(u_{h,k}(t)) \leqslant + \varepsilon'(k) + |\delta u_{h,k}(0)|^2 + a(u_{h,k}(0)) + \\ \\ + 2 \displaystyle\int_k^t (\gamma u_{h,k} + A_h u^{\theta}_{h,k}, v_{h,k})ds + 2 \int_k^t (f_{h,k}, du_{h,k} - v_{h,k})ds + \\ \\ + 2 \displaystyle\int_k^t [j_h(v_{h,k}) - j_h(du_{h,k})]\ ds \quad . \end{cases}$$

En passant à la limite on a

$$(1.83) \qquad \lim\ X_3(t) = |u'(0)|^2 + a(u(0)) + 2 \int_0^t (u'' + Au, u')ds\ ,$$

et en rassemblant les résultats, il vient finalement

$$(1.84) \qquad \lim X(t) = 0,$$

ce qui établit le théorème. ∎

Démonstration du Théorème 1.6 (suite II)

La démonstration est tout à fait semblable à celle qui précède. On définit (1.69)(1.70)(1.71). On obtient (1.72) et (1.76) sans changement.
La seule différence apparaît dans le traitement du 3ème terme. On écrira (1.77) avec $du_{h,k}$ remplacé par $\delta u_{h,k}$. On prendra $v_{h,k}$ comme en (1.78). On notera que

$$\int_{ik}^{(i+1)k} (\gamma u_{h,k} + A_h u^{\theta}_{h,k}, \delta u_{h,k})ds = (\frac{\delta^i - \delta^{i-1}}{k} + Au^{i+\theta} , \delta^i)k =$$

$$= \tfrac{1}{2} \{ |\delta^i|^2 - |\delta^{i-1}|^2 + |\delta^i - \delta^{i-1}| \} + (Au^{i+\theta} , u^{i+1} - u^i) .$$

Utilisant (1.59)(1.60)(1.61) et les majorations données par la stabilité, on obtient (1.82) avec $du_{h,k}$ remplacé par $\delta u_{h,k}$ et le résultat suit par passage à la limite. ∎

Remarque 1.3

Dans le cas où $\forall\, h,k \to 0$ l'inéquation (1.1) (resp. (1.56)) est en fait pour i=1, une équation (convexe "non saturé"), on a

$$\frac{\delta^1 - \delta^0}{k} + Au^{1+\theta} - f^0 = 0 .$$

Or

$$ku^{1+\theta} = k\theta^-(u^0 - u^1) + k\theta^+(u^2 - u^1) + ku^1 =$$

$$= -k^2\theta^- \delta^0 + ku^1 + k^2\theta^+ \delta^1 ,$$

d'où

$$\|\delta^1\| \leq \|\delta^0 + kAu^{1+\theta} - kf^0\| \leq$$

$$\leq \|\delta^0\| + k^2\theta^- \|A\delta^0\| + k\|Au^1\| + k^2\theta^+ \|A\delta^1\| +$$

$$+ k\|f^0\| ,$$

d'où l'on déduit, dès que k est assez grand, que δ^1 est borné pour la norme V_h indépendamment de h et k et l'hypothèse (1.8) est en fait inutile. ∎

2. VERIFICATION DES HYPOTHESES D'APPROXIMATION DES CONVEXES ET DES FONCTIONNELLES

2.1. Hypothèses sur K

On suppose que le triplet $\{ p_h , K_h , K \}$ vérifie les hypothèses de consistance donnée au Chapitre II et vérifiées dans ce même chapitre sur les exemples donnés. Tout d'abord les hypothèses sur l'approximation des valeurs initiales u_o et u_1 par $u_{o,h}$ et $u_{1,h}$ sont immédiates. Nous allons montrer que les hypothèses de consistance sont licites.

Lemme 2.1.

Soient $v_{h,k} \in L^2(0,T;V_h)$, $\delta_k v_{h,k}(t) \in K_h$, $t \in [0,T]$, tels que

$$v_{h,k}(0) = u_{o,h} \quad , \quad \delta_k v_{h,k}(0) = u_{1,h} \quad \text{et}$$

$$q_h v_{h,k} \longrightarrow v \quad \text{dans} \quad L^p(0,T;H) \quad \text{faible},$$

$$p_h v_{h,k} \longrightarrow v_1 \quad \text{dans} \quad L^p(0,T;F) \quad \text{faible},$$

$$q_h \delta_k v_{h,k} \longrightarrow v_2 \quad \text{dans} \quad L^p(0,T;H) \quad \text{faible},$$

$$p_h \delta_k v_{h,k} \longrightarrow v_3 \quad \text{dans} \quad L^p(0,T;F) \quad \text{faible},$$

$$q_h \gamma_k v_{h,k} \longrightarrow v_4 \quad \text{dans} \quad L^p(0,T;H) \quad \text{faible},$$

(avec $-\infty \leqslant p \leqslant 1$, et pour $p = +\infty$, la convergence étant faible étoile),

$v_1 = \sigma v$, $v_2 = v'$, $v_3 = \sigma v'$, $v'(t) \in K$, $t \in [0,T]$, $v_4 = v''$, $v(0) = u_o$, $v'(0) = u_1$.

Démonstration

Tout d'abord, d'après le Lemme 2.3 du N° 2 Ch. IV, on a $v_1 = \sigma v$, $v_3 = \sigma v_2$ et $v_2(t) \in K$, p.p. $t \in [0,T]$. Montrons que $v_2 = v'$. Tout d'abord v' est défini en tant qu'élément de $\mathcal{D}'(0,T;V)$. Pour démontrer que $v_2 = v'$ il suffit alors de montrer que $q_h \delta_k v_{h,k} \to v'$ dans $\mathcal{E}'([0,T]; H)$. Pour cela on prend $\varphi(t) \in \mathcal{C}^\infty(0,T)$ et on montre que $\forall w \in H$ on a :

$$\lim \int_0^T (q_h \delta_k v_{h,k} , \varphi(t)w)dt = - \int_0^T (v , \varphi'(t) w)dt .$$

Utilisant la densité de $\mathcal{C}^\infty(0,T) \otimes H$ dans $\mathcal{D}([0,T] ; H)$ on a alors le résultat. On fait de même pour montrer que $v_4 = v''$. Il reste à examiner que $v(0) = u_o$ et $v'(0) = u_o$. Tout d'abord comme $v \in \mathcal{C}([0,T] ; V)$ et $v' \in \mathcal{C}([0,T] ; H)$, $v(0)$ et $v'(0)$ ont un sens dans V et H respectivement. Soit $\varphi \in \mathcal{C}^1(0,T)$, $\varphi = 0$ dans un voisinage de T et soit $w \in H$. Alors[1]

$$\int_0^T (\delta_k v_{h,k}(t), \varphi(t)w)dt = \int_0^T (\frac{v_{h,k}(t+k) - v_{h,k}(t)}{k} , \varphi(t)w)dt =$$

$$= \frac{1}{k}\Big[\int_0^T (v_{h,k}(t+k) , \varphi(t)w)dt - \int_0^T (v_{h,k}(t) , \varphi(t)w)dt \Big] =$$

$$= \frac{1}{k}\Big[\int_k^T (v_{h,k}(t), \varphi(t-k)w)dt - \int_k^T (v_{h,k}(t), \varphi(t)w)dt\Big] - \frac{1}{k}\int_0^T (u_{o,h}, \varphi(t)w)dt =$$

$$= \int_k^T (v_{h,k}(t), \frac{\varphi(t-k) - \varphi(t)}{k} w)dt - \frac{1}{k} \int_0^k (u_{o,h} , \varphi(t)w)dt$$

or

$$\frac{\varphi(t-k) - \varphi(t)}{k} \to - \varphi'$$

dans $\mathcal{C}^o(0,T)$. En passant à la limite et en utilisant la formule de Green on obtient

$$(v(0) - u_o , \varphi(t)w) = 0$$

d'où, par densité, $v(0) = u_o$.

On obtiendrait de même que $v_4 = v''$ et $v'(0) = u_1$. ■

Les autres résultats de (1.6) s'obtiennent de la même façon.

---

(1) on suppose pour simplifier que θ_h = identité.

Les hypothèses de convergence sont vérifiées d'après le Lemme 2.4 , N° 2, Ch.IV.

2.2. Hypothèses sur $a(t;u,v)$

Ces hypothèses ont été vérifiées au Chapitre IV, N° 2, § 2.2.

2.3. Hypothèses sur $j(t,v)$

Les hypothèses étant sans changement par rapport à celles du Chapitre IV, ón se reportera au Chapitre IV, N° 2, § 1.3.

2.4. Hypothèses sur $f(t)$

Il faut vérifier le

Lemme 2.2.

Soit $f(t)$ tel que $f,f' \in L^\infty(0,T;H)$ alors il existe

$$f_{h,k} = \sum_{i=0}^{N} f_h^i \, \varkappa_k^i \quad , \quad f_h^i \in H \ , \ i=0,\dots,N \ \text{tel que}$$

$$\left| \begin{array}{l} q_h f_{h,k} \longrightarrow f \ \text{dans} \ L^\infty(0,T;H) \ \text{et} \ \|f_{h,k}\|_h^* \leqslant C \ , \\ q_h \delta_k f_{h,k} \longrightarrow f' \ \text{dans} \ L^\infty(0,T;H) \ \text{et} \ \|\delta_k f_{h,k}\|_h^* \leqslant C \ . \end{array} \right. \quad \square$$

Démonstration

On remarque tout d'abord que $f \in \mathcal{C}^0([0,T]; H)$ en sorte que l'on peut prendre, pour $i = 0,\dots,N$,

$$f^i = f(ik) \ ,$$

$$f_h^i = r_h f(ik) \ ,$$

où r_h a été défini au Chapitre II,§2.4. On sait que $q_h r_h f(t) \to f(t)$ dans H fort, p.p. en $t \in [0,T]$. Il nous faut montrerque $\forall \varepsilon > 0$ aussi petit que l'on veut on a

$$\sup_{t \in [0,T]} \operatorname{ess} \left| q_h f_{h,k} - f \right| \leqslant \varepsilon \qquad \text{pour } h,k \text{ assez petits}$$

$$\text{i.e.} \quad \max_{t \in [0,T]} \left| \sum_{i=0}^{N} q_h r_h f(ik) \ \varkappa_k^i - f(t) \right| \leqslant \varepsilon$$

$$\text{i.e.} \quad \max_{i=1,\dots,N} \left| q_h r_h f(ik) - f(ik) \right| + \max_{t \in [0,T]} \left| \sum_{i=1}^{N} f^i \, \varkappa_k^i - f(t) \right| \leqslant \varepsilon \ .$$

La majoration du second terme est immédiate du fait de la continuité de $f(t)$ sur $[0,T]$; quant au premier terme cela résulte du fait que

$$q_h r_h g \to g \ \text{dans} \ H \ , \quad \forall g \in H \ , \quad h \to 0$$

(cf. Ch. II,§ 2.4.) .

Il faut montrer le second point. On a, pour

$$t \in [ik,(i+1)k[\quad , \text{ et avec } f_k(t) = \sum_{i=0}^{\infty} f^i \chi_k^i(t) ,$$

$$\delta_k f_k(t) = \frac{1}{k} \int_{ik}^{(i+1)k} f'(s)ds$$

de même

$$f_h^{i+1} - f_h^i = r_h [f((i+1)k) - f(ik)]$$

$$= r_h \int_{ik}^{(i+1)k} f'(s)ds$$

d'où

$$\delta_k f_{h,k}(t) = r_h \frac{1}{k} \int_{ik}^{(i+1)k} f'(s)ds. \chi_k^i \quad .$$

Il faut alors évaluer

$$| q_h \delta_k f_{h,k}(t) - f' |_{L^\infty(0,T;H)} \leqslant | \delta_k f_{h,k} - r_h f' |_{L^\infty(0,T;H)} + | q_h r_h f' - f' |_{L^\infty(0,T}$$

Le deuxième terme du second membre tend vers 0 car $\forall g \in L^p(0,T;H)$ avec $1 \leqslant p \leqslant \infty$ on a

$$| q_h r_h g(t) | \leqslant C | g(t) | \quad \text{p.p. } t \in [0,T] \quad \text{(propriété de } q_h r_h)$$

et

$$q_h r_h g(t) \to g(t) \text{ dans } H \text{ fort , p.p. } t \in [0,T]$$

d'où, par application du Théorème de Lebesgue, $| q_h r_h g - g |_{L^\infty(0,T;H)} \to 0$.

Il reste à voir que

$$| \delta_k f_{h,k} - r_h f' |_{L^\infty(0,T;H)} \to 0$$

ou encore que

$$| \delta_k f_{h,k} - r_h \delta_k f_k |_{L^\infty(0,T;H)} + | r_h \delta_k f_k - r_h f' |_{L^\infty(0,T;H)} \to 0 \quad .$$

Comme le premier terme est nul, il reste à voir que

$$| r_h \delta_k f_k - r_h f' |_{L^\infty(0,T;H)} \to 0$$

ou encore, puisque r_h est continu, que

$$| \delta_k f_k - f' |_{L^\infty(0,T;H)} \to 0 \quad .$$

Par densité il suffit de vérifier ceci pour $f \in \mathcal{C}^1((0,T))$, ce qui est immédiat. ∎

CHAPITRE VII

RESOLUTION NUMERIQUE DE QUELQUES PROBLEMES

1. METHODES THEORIQUES.

Dans cette première partie nous étudions la résolution des inéquations de la forme

$$(Au - f, v-u) + j(v) - j(u) \geqslant 0 \quad , \qquad \forall v \in K \, , \; u \in K$$

où

K est un convexe fermé d'un espace de Banach réflexif V (de dual V')

$A \in \mathcal{L}(V,V')$, $(Av,v) \geqslant \alpha \, \|v\|_V^2$, $\forall v \in V$,

$f \in V'$,

$(,)$ est le produit scalaire de la dualité,

$j : V \to \mathbb{R}$, convexe, s.c.i.

On sait que lorsque

A est symétrique

la résolution de l'inéquation est un problème équivalent (1) au problème

trouver $u \in K$ qui

minimise $\frac{1}{2} (Av,v) - (f,v) + j(v)$.

Dans ce qui suit nous donnons des méthodes pour la résolution de ce dernier problème pour des opérateurs A symétriques, et successivement, pour $K = V = \mathbb{R}^N$, $K = \{ v \mid a_i \leqslant v_i \leqslant b_i \;\; , \;\; i=1,\ldots,N \} \subset V = \mathbb{R}^N$ et K convexe quelconque de $V=\mathbb{R}^N$.

---

(1) Si A est non symétrique il est faux, en général, que les problèmes soient équivalents. Pour la résolution des problèmes non symétriques cf. TREMOLIERES [4] .

1.1. Méthodes de minimisation sans contrainte

On rappelle brièvement les principales méthodes de minimisation sans contrainte. Le problème à résoudre est de

$$\boxed{\text{minimiser } J(v)\ ,\ v \in \mathbb{R}^N}$$

où $J(v)$ est une fonction convexe, deux fois continuement différentiable.

a) - Méthode du gradient à pas constant

On se donne $v^o \in \mathbb{R}^N$, $\rho > 0$ et l'on définit, pour n=0,1,...

$$v^{n+1} = v^n - \rho \,.\, J'(v^n)$$

Si λ_m et λ_M sont respectivement la plus petite et la plus grande valeur propre de J'' (J étant quadratique), la méthode converge si

$$0 < \rho < \rho_M = \frac{2}{\lambda_M}$$

et un choix "optimal" de ρ est $\rho_{opt} = \dfrac{2}{\lambda_m + \lambda_M}$.

b) - Méthode du gradient à pas ajusté

La méthode du gradient à pas constant n'est satisfaisante que si ρ est choisi très près de la valeur ρ optimale or ceci oblige à calculer λ_m et λ_M , ce qui est en général sinon impossible du moins très couteux. Avec une légère modification, on peut éviter le calcul de λ_m et λ_M et obtenir une méthode convergente. La méthode modifiée se présente comme suit :

On se donne $v^o \in \mathbb{R}^N$, $\rho_o > 0$, et deux constantes fixes $a > 1$, $b > 1$. On pose

$$v^n(\rho) = v^n - \rho J'(v^n) \ .$$

De façon itérative, on définit

$$v^{n+1} = v^M(\rho_{n+1}) = v^n - \rho_{n+1} \,.\, J'(v^n)$$

où ρ_{n+1} est calculé comme suit : on considère d'abord l'ensemble

$$P_h = \{ \ldots, \frac{\rho_n}{b^3}, \frac{\rho_n}{b^2}, \frac{\rho_n}{b}, \rho_n, \rho_n.a, \rho_n.a^2, \ldots \}$$

et l'on choisit ρ_{n+1} dans P_h tel que (1)

$$J(v(\rho_{n+1})) \leqslant J(v(\bar{\rho})) \quad , \quad \forall\, \bar{\rho} \in P_n \ .$$

---

(1) On peut aussi choisir $v(\rho_{n+1})$ tel que $\|J'(v(\rho_{n+1}))\| \leqslant \|J'(v(\rho))\|$, $\forall\ \rho \in P_n$.

Par cette variante la convergence est en général plus rapide et plus précise.

c) - Méthode du gradient à pas optimal

C'est la troisième variante de l'algorithme de Cauchy. On se donne $v^o \in \mathbb{R}^N$ et l'on définit

$$v^{n+1} = v^n - \rho_{n+1} J'(v^n)$$

où l'on choisit ρ_{n+1} de façon que

$$J(v^{n+1}) \leqslant J(v^n(\rho)) \ , \ \forall \ \rho \geqslant 0 \ , \text{ avec } \ v^n(\rho) = v^n - \rho J'(v^n) \ .$$

Dans le cas où $J(v)$ est quadratique, cette valeur est donnée explicitement par

$$\rho_{n+1} = \frac{\| J'(v^n) \|^2}{(J''.J'(v^n) \ , \ J'(v^n)} \ .$$

d) - Méthode du gradient conjugué

On suppose que J est quadratique. On se donne $v^o \in \mathbb{R}^N$, $r^o = -J'(u^o)$ $w^o = r^o$, et pour $n = 0,1,\ldots$ on définit

$$\left| \begin{array}{l} v^{n+1} = v^n + \mu_n w^n \quad , \quad \mu_n = \dfrac{(w^n \ , \ r^n)}{(J'' \ w^n, w^n)} \quad , \\ \\ r^{n+1} = - J'(v^{n+1}) \ , \\ w^{n+1} = r^{n+1} + \lambda_{n+1} w^n \ , \quad \lambda_{n+1} = - \dfrac{(r^{n+1}, J'' \ w^n)}{(w^n, J'' w^n)} = \dfrac{\| r^{n+1} \|^2}{\| r^n \|^2} \ . \end{array} \right.$$

Pour une forme quadratique sur $\mathbb{R}^N$, cette méthode converge en moins de N itérations. On remarquera que $J''w = J'(w) - J'(0)$ en sorte que la méthode utilise très peu de mémoires. Dans la pratique le choix le plus intéressant pour λ_{n+1} est $\dfrac{\| r^{n+1} \|^2}{\| r^n \|^2}$. Ce choix permet de programmer la méthode avec utilisation de 4 vecteurs et chaque itération demande le calcul de 3 produits scalaires : (w^n, r^n), $(J'(w^n) - J'(0), w^n)$, (r^{n+1}, r^{n+1}).

Une variante de cette méthode est la méthode du gradient conjugué à pas ajusté ; cette variante a son intérêt dans le cas où J est non quadratique. On pose

$$v^{n+1} = v^n + \rho_{n+1} w^n$$

où ρ_{n+1} est choisi comme en b). Puis l'on pose $r^{n+1} = - J'(v^{n+1})$ et

$$\tilde{w}^{n+1} = r^{n+1} + \lambda_{n+1} w^n \ , \quad \lambda_{n+1} = \| r^{n+1} \|^2 / \| r^n \|^2 \ .$$

On définit alors

$$w^{n+1} = \left\{ \begin{array}{l} \tilde{w}^{n+1} \quad \text{si } (\tilde{w}^{n+1}, r^{n+1}) \geqslant \varepsilon > 0, \ \varepsilon \text{ constante fixe } < 1 \ , \\ \\ r^{n+1} \quad \text{sinon .} \end{array} \right.$$

Une autre variante (dans le cas non quadratique) consiste à prendre ρ_{n+1} comme en c).

e) - Méthode de Fletcher-Powell

Nous renvoyons à FLETCHER et POWELL [1], CEA [2] pour la description de cet algorithme. Comme pour la méthode du gradient conjugué, cette méthode converge en moins de N itérations. Elle utilise 4 vecteurs et une matrice. De plus chaque itération demande des produits matriciels. ■

f) - Méthode de relaxation (ou de Gauss-Seidel relaxée)

On se donne $v^o \in \mathbb{R}^N$ et l'on calcule itérativement, pour $n=0,\ldots,v^{n+1}$ composante par composante, i.e., n étant fixé et v^n connu, on définit, pour $i=1,2,\ldots,N$,

$$v_i^{n+\frac{1}{2}} = v_i^n + \rho_i^{n+\frac{1}{2}}$$

où $\rho_i^{n+\frac{1}{2}}$ est calculé de telle façon que

$$J'(v_1^{n+1},\ldots, v_{i-1}^{n+1}, v_i^{n+\frac{1}{2}}, v_{i+1}^n,\ldots, v_N^n) = 0$$

i.e.

$$J(v_1^{n+1},\ldots, v_{i-1}^{n+1}, v_i^n + \rho_i^{n+\frac{1}{2}}, v_{i+1}^n,\ldots, v_N^n) \leqslant$$

$$\leqslant J(v_1^{n+1},\ldots,v_{i-1}^{n+1}, v_i^n + \rho, v_{i+1}^n,\ldots v_N^n), \quad \forall \rho \in \mathbb{R}$$

puis l'on pose

$$v_i^{n+1} = v_i^n + \omega (v_i^{n+\frac{1}{2}} - v_i^n) .$$

La méthode est convergente pour $0<\omega<2$. Dans le cas quadratique $\rho_i^{n+\frac{1}{2}}$ est donné par (1)

$$\rho_i^{n+\frac{1}{2}} = - \frac{J_i'(v^{n(i)})}{a_i^i} , \quad \text{avec} \quad J'' = [a_i^j]$$

et

$$v^{n(i)} = (v_1^{n+1},\ldots, v_{i-1}^{n+1}, v_i^n, v_{i+1}^n,\ldots, v_N^n)$$

Toujours dans le cas quadratique, une valeur asymptotiquement "optimale" du paramètre ω est

$$\omega_{opt} = \frac{2}{1+\sqrt{1-\rho(\mathcal{L}_1)}}$$

où $\rho(\mathcal{L}_1)$ est le rayon spectral (i.e. la plus grande valeur propre) de la matrice dite de Gauss-Seidel :

$$\mathcal{L}_1 = -(D+L)^{-1}U$$

---

(1) On définit $v'=v+\rho d$ et $J(v') \leqslant J(v+\rho d)$, $\forall \rho$ $\Leftrightarrow$ (dans le cas quadratique convexe) $0 = J'(v')d = J'(v)d + \rho(J''d,d)$.

avec $J'' = D + L + U$,

D = matrice diagonale extraite de J'' ,

L = matrice triangulaire inférieure extraite de J'',

U = matrice triangulaire supérieure extraite de J'' .

En supposant que $J(v)$ s'écrit

$$J(v) = \tfrac{1}{2}(Av,v) + (b,v) + c \quad .$$

une façon de déterminer ω_{opt} est de résoudre $Av = 0$ par le schéma

$$u^{n+1} = \mathcal{L}_1 u^n \quad , \quad \text{avec} \quad u^o > 0 \quad \text{donné}$$

et, posant

$$\rho_p^{n+1} = \min_i \left(\frac{u_i^{n+1}}{u_i^n}\right) \quad , \qquad \rho_g^{n+1} = \max_i \left(\frac{u_i^{n+1}}{u_i^n}\right) \quad ,$$

$$\omega_p^{n+1} = \frac{2}{1 - \sqrt{1 - \rho_p^{n+1}}} \quad , \qquad \omega_g^{n+1} = \frac{2}{1 - \sqrt{1 - \rho_g^{n+1}}} \quad ,$$

on peut montrer que

$$\rho_p^1 \leqslant \rho_p^2 \leqslant \cdots \leqslant \rho_p^n \leqslant \cdots, \cdots \leqslant \rho_g^n \leqslant \cdots \leqslant \rho_g^2 \leqslant \rho_g^1 \quad ,$$

$$\omega_p^1 \geqslant \omega_p^2 \geqslant \cdots \geqslant \omega_p^n \geqslant \cdots \rho(\mathcal{L}_1) \cdots \geqslant \omega_g^n \geqslant \cdots \geqslant \omega_p^2 \geqslant \omega_p^1 \quad ,$$

et que

$$\rho_p^n - \rho_g^n \to 0 \quad , \qquad \omega_p^n - \omega_g^n \to 0 \quad ,$$

$$\omega_p^n \to \rho(\mathcal{L}_1) \quad ,$$

$$\omega_g^n \to \rho(\mathcal{L}_1) \quad .$$

En particulier

$$\omega_{opt}^n = \frac{\omega_p^n + \omega_g^n}{2} \to \rho(\mathcal{L}_1) \quad . \qquad \blacksquare$$

1.2. Méthodes de minimisation avec contraintes de bornes

Toutes les méthodes décrites précédemment peuvent s'adapter sans difficulté à la résolution des problèmes avec contraintes de bornes, i.e.

$$\left|\begin{array}{l} \underline{\text{minimiser}} \quad J(v) \quad , \quad v \in \mathbb{R}^N \\ v \in K \quad \text{avec} \\ K = \prod_{i=1}^{N} [a_i, b_i] \quad , \quad a_i \leqslant b_i \in \bar{\mathbb{R}} \quad , \quad i = 1, \ldots, N \quad . \end{array}\right.$$

Notant P_k la projection au sens usuel dans $\mathbb{R}^N$ les méthodes itératives a)b)c) d)e)f) peuvent être modifiées comme suit :

a) - 1ère manière : on définit

$$v^{n+1} = P_k \left[v^n - \rho J'(v^n) \right] .$$

- 2ème manière : on remplace $v^{n+1} = v^n - \rho J'(v^n)$ par

$$v^{n+1} = v^n - \rho^n d^n$$

avec

$$d_i^n = J_i'(v^n) \quad \text{si} \quad a_i < v_i^n < b_i ,$$

$$d_i^n = J_i'(v^n) \quad \text{si} \quad v_i^n = a_i \text{ et si } -J_i'(v^n) \geqslant 0 ,$$

$$d_i^n = J_i'(v^n) \quad \text{si} \quad v_i^n = b_i \text{ et si } -J_i'(v^n) \leqslant 0 ,$$

$$d_i^n = 0 \quad \text{sinon} .$$

On définit alors v^{n+1}, ρ^n par

$$v^{n+1} = v^n - \rho^n d^n \quad \text{si} \quad v^{n+1} \in K \quad , \quad (\rho^n = \rho')$$

sinon

$$v^{n+1} = v^n - \rho^n d^n$$

où ρ^n est la plus grande valeur de $\rho \geqslant 0$ telle que $v^{n+1} \in K$. ∎

b') - c') - Pour les méthodes b) et c) on peut procéder comme ci-dessus.

d') - e') - On peut procéder comme en a) 2ème manière en remplaçant dans a'), $J_i'(v^n)$ par la direction (conjuguée) w^n. Il faut cependant vérifier que la direction d^n a une projection positive sur $J'(v^n)$, sinon, on remplace w^n par $- J'(v^n)$.

f') - On remplace la formule de relaxation de f) par

$$v_i^{n+1} = P_k \left[v_i^n + \omega (v_i^{n+\frac{1}{2}} - v_i^n) \right] . \quad ∎$$

Devant l'intérêt d'une méthode particulière dérivée de celle du gradient conjugué nous donnons cette méthode dans ce qui suit :

Méthode S.N.

On se propose de

minimiser $J(v) = \frac{1}{2}(Av,v) - (f,v)$
avec $v=(v_1,\ldots,v_N) \in \mathbb{R}^N$ sous les contraintes
$a_i \leqslant v_i \leqslant b_i$, $a_i , b_i \in \bar{\mathbb{R}}$, $i=1,\ldots N$.

(Nous appelons K le sous-ensemble de $\mathbb{R}^N$ défini par ces contraintes

Avant d'exposer la méthode, nous précisons d'abord deux termes :

- composante bloquée,
- pas "optimal".

A l'itération n, soit v^n un vecteur connu. Nous disons qu'une composante v_i^n de v^n est bloquée si

$$v_i^n = a_i \quad \text{et si} \quad J_i'(v^n) \leqslant 0$$

ou bien si

$$v_i^n = b_i \quad \text{et si} \quad J_i'(v^n) \geqslant 0 \ .$$

Pour la commodité nous définissons un vecteur de repérage des contraintes bloquée à l'itération n et relativement au vecteur v^n :

$$\delta^n = (\delta_1^n, \ldots, \delta_N^n)$$

avec

$$\delta_i^n = \begin{cases} 0 \text{ si } v_i^n \text{ est bloqué ,} \\ 1 \text{ sinon.} \end{cases}$$

Etant donné un vecteur w nous notons

$$\delta^n . w = (\delta_1^n . w_1, \ldots, \delta_N^n . w_N)$$

Par ailleurs étant donné v^n un vecteur connu et w un autre vecteur, nous définissons ρ^n comme étant le pas "optimal" dans la direction w si

$$J(v^n - \rho^n w) \leqslant J(v^n - \rho . w)$$

ceci $\forall \rho \in \mathbb{R}$ tel que $v^n - \rho . w \in K$

et ρ^n étant tel que $v^n - \rho^n . w \in K$.

L'algorithme procède alors comme suit :

1ère itération

Etant donné $v^o \in K$, définir le vecteur des composantes bloquées de v^o, soit δ^o, puis déterminer

$$v^1 = v^o + \rho^o . d^o$$

où ρ^o est le pas optimal dans la direction d^o avec :

$$d^o = - \delta^o . J'(v^o) \ .$$

$(n+1)^e$ itération $(n \geqslant 1)$

Etant donné v^n, déterminer $J'(v^n)$ et δ^n puis calculer :

$$v^{n+1} = v^n + \rho^n d^n$$

où ρ^n est le pas optimal dans la direction d^n, cette direction étant déterminée comme suit :

si $\delta^n = \delta^{n-1}$, poser, pour $i=1,\dots,N$,

$d_i^n = -\delta_i^n \cdot J_i'(v^n) + \lambda_n . d_i^{n-1}$, $\lambda_n = \dfrac{\| \delta^n J'(u^n) \|^2}{\| \delta^{n-1} J'(v^{n-1}) \|^2}$;

si $\forall i$ tel que $v_i^n = a_i$ on a $d_i^n \leqslant 0$

et si $\forall i$ tel que $v_i^n = b_i$ on a $d_i^n \geqslant 0$

poser

$$d^n = (d_1^n, \dots, d_N^n)$$

sinon poser

$$d^n = -\delta^n . J_i'(v^n) ,$$

si $\delta^n \neq \delta^{n-1}$, poser

$$d^n = -\delta^n . J_i'(v^n) . \quad \blacksquare$$

Dans son principe cet algorithme peut être rapproché des considérations établies dans TREMOLIERES [6] . ■

1.3. Méthodes de minimisation sous contraintes convexes

1.3.1. Méthode duale semi-implicite

On cherche à résoudre le

Problème P. (primal)

$$\begin{cases} \min J(v) \\ G(v) \in D \\ v \in C \end{cases}$$

où

$v \in \mathbb{R}^N$,

$J(v) \in \mathcal{C}^1(\mathbb{R}^N, \mathbb{R})$,

$(J'(v) - J'(w), v-w) \geqslant \alpha \| v-w \|^2$ ($\|.\|$ désigne la norme sur $\mathbb{R}^N$ et $(,)$ le produit scalaire sur $\mathbb{R}^N$) ,

$G(v) : \mathbb{R}^N \to \mathbb{R}^M$, $M \geqslant 1$,

$| G(v) - G(w) | \leqslant S \| v-w \|$ ($|.|$ désigne la norme sur $\mathbb{R}^M$ et S est

une constante >0) ,

C est un ensemble convexe fermé de $\mathbb{R}^N$,

D est un cône convexe fermé de $\mathbb{R}^M$ de sommet l'origine (1),

On définit le cône dual de D par

$$D^* = \{ q \mid q \in \mathbb{R}^M , \ qs \leq 0 , \ \forall s \in D \} .$$

Le Lagrangien $L(v,q)$ du Problème P est défini par l'expression :

$$L(v,q) = J(v) + (q, G(v)) \qquad v \in \mathbb{R}^N , \ q \in \mathbb{R}^M .$$

Rappelons que si $(u,p$ est point selle de $L(v,q)$ sur $C \times D^*$, i.e.

$$L(u,q) \leq L(u,p) \leq L(v,p) , \quad \forall (v,q) \in C \times D^* , \ (u,p) \in C \times D^* ,$$

alors u est solution du Problème P. De plus, une condition nécessaire et suffisante pour que (u,p) soit point-selle de $L(v,q)$ sur $C \times D^*$ (ceci sous des hypothèse que l'on suppose vérifiées ici (2)) est que

$$(1.1) \qquad (J'(u), w-u) + (p, G(w) - G(u)) \geq 0 , \quad \forall w \in C , \ u \in C ,$$

$$(1.2) \qquad - (G(u), q-p) \geq 0 , \quad \forall q \in D^* , \ p \in D^* \qquad \blacksquare$$

Algorithme

On se donne $q^0 \in D^*$ et pour $n=0,1,\ldots$ on détermine successivement $v^{n+1} \in C$ et $q^{n+1} \in D^*$ tels que

$$(1.3) \qquad (J'(v^{n+1}), w-v^{n+1}) + (q^n, G(u) - G(v^{n+1})) \geq 0 , \ \forall w \in C, \ v^{n+1} \in C , \ (3)$$

$$(1.4) \qquad (q^{n+1} - q^n - \rho G(v^{n+1}) , r-q^{n+1}) \geq 0 , \ \forall r \in D^* , \ q^{n+1} \in D^* ,$$

avec

$$(1.5) \qquad 0 < \varepsilon \leq \rho \leq \frac{2\alpha}{S^2} - \varepsilon , \quad \text{où} \quad 0 < \varepsilon \leq \frac{\alpha}{S^2} . \ \blacksquare$$

Nous allons démontrer le

Théorème 1.1. Avec $n \to +\infty$, on a

. $v^n \to u$,

. q^n reste borné,

---

(1) Par exemple $D = \{ s \mid s \in \mathbb{R}^M , \ s = (s_1 , \ldots , s_M) , \text{ avec } s_i \leq 0, \ i=1,\ldots,m \leq M$ et(si $m<M$) $s_k = 0$, $k=m+1,\ldots,M$. Le Problème P peut donc comporter des "contraintes" en inégalité aussi bien que des contraintes d'égalité.

(2) cf. TREMOLIERES [2] [7] [8] .

(3) i.e. $L(v^{n+1}, q^n) \leq L(v, q^n)$, $\forall v \in C$, $v^{n+1} \in C$.

$\cdot\ |q^{n+1} - q^n| \to 0$,

$\cdot$ si p est une valeur d'adhérence de q^n alors (u,p) est point-selle de $L(v,q)$ sur $C \times D^*$. ∎

Démonstration

Montrons d'abord les deux premiers points. Faisons $w=v^{n+1}$ dans (1.1)' et $w=u$ dans (1.3). En ajoutant, il vient

$$(1.6) \qquad (J'(v^{n+1}) - J'(u)\ ,\ v^{n+1} - u) + (q^n - p,\ G(v^{n+1}) - G(u)) \leqslant 0 \ .$$

Multiplions (1.2) par ρ et remplaçons q par q^{n+1} dans (1.2) et r par p dans (1.4) ; en ajoutant, il vient

$$(1.7) \qquad (q^{n+1} - p - [q^n - p] - \rho[G(v^{n+1}) - G(u)]\ ,\ p - q^{n+1}) \geqslant 0$$

d'où

$$|q^{n+1} - p|^2 \leqslant (q^n - p + \rho\ [\ G(v^{n+1}) - G(u)]\ ,\ q^{n+1} - p) \ ,$$

$$|q^{n+1} - p| \leqslant |q^n - p + \rho\ [\ G(v^{n+1}) - G(u)]\ | \ .$$

En élevant au carré et en développant, il vient

$$(1.8) \qquad |q^{n+1} - p|^2 \leqslant |q^n - p|^2 + 2\,\rho(q^n - p\ ,\ G(v^{n+1}) - G(u)) + \rho^2\,|G(v^{n+1}) - G(u)|^2 \ .$$

Les deux inégalités (1.6) et (1.8) donnent

$$2\rho\alpha\,\|v^{n+1} - u\|^2 + |q^{n+1} - p|^2 - |q^n - p|^2 - \rho^2\,|G(v^{n+1}) - G(u)|^2 \leqslant 0 \ .$$

Comme G est Lipschitzien, on obtient

$$(2\rho\alpha - \rho^2 S^2)\,\|v^{n+1} - u\|^2 + |q^{n+1} - p|^2 - |q^n - p|^2 \leqslant 0 \ .$$

Par sommation il vient

$$(2\rho\alpha - \rho^2 S^2) \sum_{i=0}^{n} \|v^{i+1} - u\|^2 + |q^{n+1} - p|^2 - |q^0 - p|^2 \leqslant 0 \ .$$

Mais d'après (1.5) il vient

$$(2\rho\alpha - \rho^2 S^2) > 0 \ .$$

Donc $\|v^{n+1} - u\|^2 \to 0$ et q^n reste borné. ∎

Montrons le troisième point. Il est clair que (1.4) peut s'écrire

$$(q_j^{n+1} - q_j^n - \rho\, G_j(v^{n+1})\ ,\ r_j - q_j^{n+1}) \geqslant 0\ ,\ \forall j=1,\ldots,M \ ,$$

avec

$$q^{n+1} = (q_1^{n+1},\ldots,\ q_j^{n+1},\ldots,\ q_M^{n+1}) \ ,\ q^n = (q_1^n,\ldots,\ q_j^n,\ldots,\ q_M^n) \ ,$$

$$G(v^{n+1}) = (G_1(v^{n+1}), \ldots, G_j(v^{n+1}), \ldots, G_n(v^{n+1})) ,$$

$$r = (r_1 \ \ldots, r_j, \ldots r_n) ,$$

et ceci $\forall\, r_j$ tel que $r = (0,\ldots,0,r_j,0,\ldots,0) \in D^*$.
Faisons $r_j = q_j^n$. Alors

$$|q_j^{n+1} - q_j^n|^2 \leqslant \rho(G_j(v^{n+1}),\ q_j^{n+1} - q_j^n) \quad . \tag{1.9}$$

En faisant successivement $r_j = 2q_j^{n+1}$ et $r_j = \frac{1}{2}\, q_j^{n+1}$ on obtient aussi

$$(q_j^{n+1} - q_j^n - \rho\ G_j(v^{n+1}),\ q_j^{n+1}) = 0 \quad .$$

Soit

$$|q_j^{n+1}|^2 - |q_j^n|^2 + |q_j^{n+1} - q_j^n|^2 = 2\rho\,(G_j(v^{n+1}),\ q_j^{n+1}) \quad . \tag{1.10}$$

Deux cas sont à considérer

1) - $G_j(u) = 0$. Alors $G_j(v^{n+1}) \to 0$ et (1.14) donne

$$|q_j^{n+1} - q_j^n| \to 0 \quad .$$

2) - $G_j(u) \neq 0$. (Supposons par exemple $G_j(u) < 0$).

Alors $v^n \to u$ et G_j continu entraînent

$\exists\, n_o$ tel que $G_j(v^n) < 0$, $\forall\, n \geqslant n_o$, n_o entier assez grand.

Donc $q_j^n\ G_j(v^n) \leqslant 0$. En sommant (1.10) de n_o à $n, n > n_o$, il vient

$$|q_j^{n+1}|^2 + \sum_{i=n_o}^{n} |q_j^{i+1} - q_j^i|^2 \leqslant |q_j^{n_o}|^2 \quad .$$

D'où

$$|q_j^{n+1} - q_j^n| \to 0 \quad .$$

Ce qui établit le troisième point. ∎

Il reste à montrer le dernier point. Avec (1.4) on a

$$(q^{n+1} - q^n\ ,\ q - q^{n+1}) \geqslant \rho\,(G(v^{n+1}),\ q - q^{n+1}), \quad \forall\ q \in D^* \ .$$

D'après le point précédent et le fait que q^n reste borné il vient

$$\lim_{n \to \infty} (G(v^{n+1})\ ,\ q - q^{n+1}) \leqslant 0 \quad , \quad \forall q \in D^* \quad .$$

Mais $v^{n+1} \to u$ et q^{n+1} reste borné. On peut donc extraire de q^n une sous-suite $q^{n'}$ telle que

$$q^{n'} \to \bar{p} \in D^* ,$$

$$(G(u), q-\bar{p}) \leq 0 , \quad \forall\, q \in D^* . \tag{1.11}$$

Il reste à montrer que $(u,\bar{p})$ est point-selle de $L(v,q)$ sur $C\times D^*$. Par définition de v^{n+1} on a

$$J(v^{n+1}) + (q^n, G(v^{n+1})) \leq J(v) + (q^n, G(v)) , \quad \forall v \in C .$$

Comme $v^n \to u$ et $q^{n'} \to \bar{p}$ on a

$$J(u) + (\bar{p}, G(u)) \leq J(v) + (\bar{p}, G(v)) , \quad \forall v \in C . \tag{1.12}$$

Par ailleurs, en utilisant (1.11), il vient

$$J(u) + (q, G(u)) \leq J(u) + (\bar{p}, G(u)) . \tag{1.13}$$

Les inégalités (1.12) et (1.13) montrent que $(u,\bar{p})$ est point-selle de $L(v,q)$ sur $C\times D^*$ et ceci termine la démonstration du Théorème 1.1. ∎

1.3.2. Méthode duale de double gradient

Le problème à résoudre étant toujours le Problème P, nous conservons les mêmes hypothèses que celles du § 1.1. et nous y ajoutons les suivantes :

$$\left|\begin{array}{l} G(v) = Bv-b , \\ B \in \mathcal{L}(\mathbb{R}^N , \mathbb{R}^M) , \quad b \in \mathbb{R}^M , \\ \| J'(v) - J'(w) \| \leq C \| v-w \| , \quad C \text{ constante positive} . \end{array}\right.$$

Un point-selle (u,p) de $L(v,q)$ sur $C\times D^*$ est caractérisé par :

$$(J'(u), w-u) + (B^*p, w-u) \geq 0 , \quad \forall w \in C , \quad u \in C , \tag{2.1}$$

$$(-G(u), r-p) \geq 0 , \quad \forall\, r \in D^* , \quad p \in D^* . \tag{2.2}$$

Algorithme

On se donne $(v^0,q^0) \in C\times D^*$ et pour $n=0,1,\ldots$ on détermine successivement $v^{n+1} \in C$ et $q^{n+1} \in D^*$ tels que

$$(v^{n+1} - v^n + \mu (J'(v^n) + B^*q^n , w-v^{n+1}) \geq 0 , \ \forall\, w \in C , \ v^{n+1} \in C , \tag{2.3}$$

$$(q^{n+1} - q^n - \rho G(v^{n+1}), r-q^{n+1}) \geq 0 , \ \forall\, r \in D^* , \ q^{n+1} \in D^* \tag{2.4}$$

où ρ et μ sont deux scalaires positives qui seront précisés ultérieurement. ∎

Nous allons démontrer le

Théorème 1.2. On a

- $v^n \to u$,
- q^n reste borné ,
- $|q^{n+1} - q^n| \to 0$,
- si p est une valeur d'adhérence de q^n alors (u,p) est point-sell de $L(v,q)$ sur $C \times D^*$. ∎

Démonstration

Multiplions (2.1) par μ et faisons $w=v^{n+1}$ dans (2.1) et $w=u$ dans (2.3) En ajoutant, il vient

$$(v^{n+1} - u - [v^n - u] + \mu[J'(v^n) - J'(u) + B^*q^n - B^*p] \ , \ u-v^{n+1}) \geqslant 0$$

$$u \in C \ , \ v^{n+1} \in C \ .$$

Soit

$$(2.5) \quad \|v^{n+1} - u\|^2 \leqslant (1 + \mu C) \|v^n - u\| \ \|v^{n+1} - u\| + \mu (q^n - p \ , \ Bu - Bv^{n+1}) \ .$$

En multipliant (2.2) par ρ et en faisant $r = q^{n+1}$ dans (2.2) et $r=p$ dans (2.4) il vient (comme pour (1.81))

$$(2.6) \quad |q^{n+1} - p|^2 \leqslant |q^n - p|^2 + 2\rho (q^n - p, Bv^{n+1} - Bu) + \rho^2 |Bv^{n+1} - Bu|^2 \ .$$

Les deux inégalités (2.5)(2.6) donnent (avec $\mu = \frac{2\rho}{\beta}$, $\beta > 0$)

$$|q^{n+1} - p|^2 \leqslant |q^n - p|^2 + \beta (1 + \mu C) \|v^n - u\| . \|v^{n+1} - u\| - \beta \|v^{n+1} - u\|^2 + \rho^2 S^2 \|v^{n+1} - u\|^2 \ .$$

Or

$$\|v^n - u\| \ \|v^{n+1} - u\| \leqslant \tfrac{1}{2}(\|v^n - u\|^2 + \|v^{n+1} - u\|^2) \ .$$

D'où

$$|q^{n+1} - p|^2 \leqslant |q^n - p|^2 + \frac{\beta}{2}(1+ \mu C) \|v^n - u\|^2 + \frac{\beta}{2}(1+ \mu C) \|v^{n+1} - u\|^2 - \beta \|v^{n+1} - u\|^2 + (\rho^2 S^2 - \beta) \|v^{n+1} - u\|^2 \ .$$

En sorte que

$$|q^{n+1} - p|^2 \leqslant |q^n - p|^2 + \frac{\beta}{2}(1+\mu C) \|v^n - u\|^2 - \frac{\beta}{2}(1+ \mu C) \|v^{n+1} - u\|^2 + \beta \mu C \|v^{n+1} - u\|^2 + (\rho^2 S^2 - \beta) \|v^{n+1} - u\|^2 \ .$$

Donc

$$|q^n-p|^2 + \frac{\beta}{2}(1+\mu C)\,\|v^n - u\|^2 - \left(|q^{n+1} - p|^2 + \frac{\beta}{2}(1+\mu C)\,\|v^{n+1} - u\|^2\right) \geqslant$$

$$\geqslant \left(\beta - \beta\mu C - \rho^2 S^2\right)\,\|v^{n+1} - u\|^2 .$$

En choisissant ρ et β (donc μ) de façon que

$$\beta\,(1-\mu C) - \rho^2 S^2 \geqslant \gamma > 0$$

il résulte que

$$\tau^n = |q^n - p|^2 + \frac{\beta}{2}(1+\mu C)\,\|v^n - u\|^2$$

décroit, donc converge et alors $\tau^n - \tau^{n+1} \to 0$ d'où il résulte que $\|v^{n+1} - u\| \to 0$ et q^n reste borné. ∎

Pour les points 3 et 4 du Théorème 1.2 la démonstration s'établit exactement de la même façon que pour le Théorème 1.1 , **exceptée** l'inégalité (1.12) que l'on obtient en passant à la limite dans l'inéquation (2.3) . On obtient

$$(J'(u),\ w-u) + (B^*\bar{p}\ ,\ w-u) \geqslant 0\ , \quad \forall w \in C,\ u \in C$$

et d'après les hypothèses faites ceci est équivalent à

$$J(u) + (\bar{p}, G(u)) \leqslant J(v) + (\bar{p}, G(v))\ , \quad \forall v \in C\ .$$

Le Théorème 1.2 est donc démontré. ∎

Remarque 1.1.

Dans le cas non linéaire, c'est à dire lorsque $G(v) \in \mathcal{C}^1(\mathbb{R}^N, \mathbb{R}^M)$ l'algorithme s'écrit :

$$\begin{cases} (v^{n+1} - v^n + \mu\,[J'(v^n) + q^n G'(v^n)]\ ,\ w-v^{n+1}) \geqslant 0\ ,\ \forall\, w \in C\ ,\ v^{n+1} \in C\ , \\ (q^{n+1} - q^n - \rho\, G(v^{n+1})\ ,\ r-q^{n+1}) \geqslant 0\ ,\ q^{n+1} \in D^*\ , \quad \forall r \in D^*\ . \end{cases}$$ ∎

Remarque 1.2.

L'inéquation (2.3) définit v^{n+1} par un calcul explicite. On peut aussi envisager d'obtenir v^{n+1} de façon implicite, par exemple en remplaçant (2.3) par

$$(v^{n+1} - v^n + \mu\,[J'(v^{n+1}) + B^*q^n]\ ,\ w-v^{n+1}) \geqslant 0\ ,\ v^{n+1} \in C,\ \forall w \in C\ .$$

On pourra reprendre la démonstration. On constatera que le premier point peut s'établir plus aisément en utilisant la monotonie de J'. ∎

Remarque 1.3.

En dimension "infinie", on prémultipliera $Au^n - f$ par $\Lambda \in \mathcal{L}(V',V)$, Λ coercif, continu. ■

1.4. Méthodes de pénalisation par troncature.

On cherche à résoudre le Problème P écrit comme suit

$$\left|\begin{array}{l} \underline{\text{minimiser}}\ \lambda \\ G_i(v) \geqslant 0 \quad , \quad i=1,\ldots,M \\ J(v) \leqslant \lambda \\ \lambda \in \mathbb{R} \end{array}\right.$$

Le paramètre λ est appelé paramètre de troncature. Soit

$$K = \{\, v \mid G_i(v) \geqslant 0 \quad , \quad i=1,\ldots,M \,\} \ .$$

Sur le domaine

$$K(\lambda) = K \cap \{\quad v \mid J(v) \leqslant \lambda \,\} \quad , \quad (\lambda > J(u)) \ .$$

On définit une fonction-distance qui, à λ fixé, permet d'associer un point aussi "central" que possible dans $K(\lambda)$. Donnons trois exemples de telles fonctions

$$D_p(v,\lambda) = (\lambda - J(v))^p \prod_{i=1}^{N} G_i(v) \quad , \quad p \text{ scalaire} \geqslant \varepsilon > 0 \quad ,$$

$$F(v,\lambda) = \frac{1}{\lambda - J(v)} + \sum_{i=1}^{N} \frac{1}{G_i(v)} \quad ,$$

$$H(v,\lambda) = \min\{\, G_1(v),\ldots,G_M(v) \,\} \ .$$

La fonction $D_p(v,\lambda)$ a les particularités suivantes :

$$\left|\begin{array}{l} = 0 \text{ si } J(v) = \lambda \quad \text{ou si} \quad \exists i \ , \ 1 \leqslant i \leqslant M \ , \text{ tel que } G_i(v) = 0 \\ > 0 \text{ si } J(v) < \lambda \quad \text{et si} \quad G_i(v) > 0 \ , \ \forall_i \ , \ 1 \leqslant i \leqslant M \quad (1) \\ \\ \text{intuitivement d'autant plus grandes que } v \text{ est plus intérieur à} \\ K(\lambda). \end{array}\right.$$

Un point $v(\lambda) \in K(\lambda)$ tel que :

$$D_p(v(\lambda),\lambda) = \max_{v \in K(\lambda)} D_p(v,\lambda)$$

---

(1) On suppose que $\overset{\circ}{K} = \emptyset$ et, $\overset{\circ}{K} \equiv \{\, v \mid G_i(v) > 0 \ , \ i=1,\ldots,M \,\}$.

ou

$$H(v(\lambda),\lambda) = \max_{v\in K(\lambda)} H(v,\lambda)$$

ou

$$F(v(\lambda),\lambda) = \min_{v\in K(\lambda)} F(v,\lambda)$$

est appelé "centre" du domaine $K(\lambda)$. Le principe des méthodes par troncature consiste alors à déterminer une suite de "centres".

$$v_1 = v(\lambda_1),\ v_2 = v(\lambda_2),\ \ldots\ v_n = v(\lambda_n),\ \ldots$$

correspondant à une suite décroissante λ_1 , λ_2 ,..., λ_n,... de troncatures. En prenant pour fonction de pénalisation la fonction $D(v,\lambda)$ l'algorithme procède comme suit :

1.4.1. Algorithme

On se donne $v_o \in K$ et l'on pose $\lambda_1 = J(v_o)$, puis, pour n=1,2,..., on définit

$v_n \in K(\lambda_n)$ tel que $D(v_n,\lambda_n) \geqslant D(v,\lambda_n)$, $\forall v\in K(\lambda_n), v_n \in K(\lambda)$,

$w_n \in K(\lambda_n)$ quelconque tel que $J(w_n) \leqslant J(v_n)$,

ρ_n scalaire quelconque choisi dans un intervalle fixe $[\delta, 1]\delta\in]0,1[$

$\lambda_{n+1} = \lambda_n - \rho_n\ [\lambda_n - J(w_n)]$.

En général on prend ρ_n = constante indépendante de n et p (dans $D_p(v,\lambda)=(\lambda - J(v))^p \prod_{2=1}^{N} G_i(v)$) variable avec n (en général (p_n) sera une suit décroissant vers 1).

1.4.2. Convergence

On a le

Théorème 1.5.

Si $K = \{ v \mid G_i(v) \geqslant 0 ,\ i=1,..,M \}$ est un convexe fermé borné d'intérieur non vide défini par des contraintes $G_i(v)$, i=1,...,M, convexes et si $J(v)$ est une fonction continue, strictement convexe alors les suites v_n et w_n données par l'algorithme vérifient :

. $v_n \longrightarrow u$ et $J(v_n) \longrightarrow J(u)$,

. $w_n \longrightarrow u$ et $J(w_n) \longrightarrow J(u)$. ∎

Nous renvoyons à TREMOLIERES [5] et GLOWINSKI-LIONS-TREMOLIERES [1] pour la démonstration de ce Théorème. Dans le cas ou $\rho_n = 1$, $w_n = v_n$, $p_n = p$ fixe, la méthode est la méthode des centres de HUARD [1].

Remarques 1.4.

Le choix de w_n peut se faire de multiples façons. Tout d'abord, on peut prendre $w_n = v_n$. Alors, si $\rho_n = 1$, la méthode correspondante (avec $p_n = p$ fixe) est la méthode des centres. Si $\rho_n < 1$, la méthode peut être considérée comme une méthode de sous-relaxation. A la limite si ρ_n est assez petit et si l'on prend $w_n = v_n$, la suite v_n décrit une courbe dont une extrêmité est justement la solution u cherchée. En général, la convergence est beaucoup plus rapide lorsque l'on prend $w_n \neq v_n$ et w_n "au mieux". Nous allons préciser deux choix possibles pour la détermination de w_n. Ce choix consistent à "extrapoler" v_n, en cherchant un point w_n meilleur pour $J(v)$, i.e. $J(w_n) < J(v_n)$. ■

Un premier procédé est le suivant :

notant $$[v_n , J'(v_n)] = \{ v \mid v \in K,\ v = v_n - \rho_n J'(v_n),\ \rho \geqslant 0 \},$$

on choisit $w_n \in [v_n, J'(v_n)]$ tel que

$$J(w_n) \leqslant J(w), \qquad \forall w \in [v_n , J'(v_n)]. \ ■$$

Un second procédé consiste en ceci :

notant $$[v_{n-1} , v_n] = \{ v \mid v \in K,\ v = v_{n-1} + \rho [v_n - v_{n-1}],\ \rho \geqslant 0 \},$$

on choisit $w_n \in [v_{n-1} , v_n]$ tel que

$$J(w_n) \leqslant J(w), \qquad \forall w \in [v_{n-1} , v_n]. \ ■$$

Avec w_n choisi tel que $J(w_n) < J(v_n)$ et ρ_n tel que

$$J(v_n) > \lambda_{n+1} \geqslant J(w_n)$$

la méthode peut être considérée comme une méthode de sur-relaxation. ■

2. EXEMPLES NUMERIQUES

Cette dernière partie est consacrée à la résolution de quelques problèmes de la physique des milieux continus et à la comparaison de quelques méthodes sur des exemples simplifiés.

Nous examinerons successivement :

- le problème de Dirichlet (2.1),
- le problème de Neumann (2.2),
- un problème de climatisation stationnaire (2.3),
- un problème non différentiable stationnaire (2.4),
- un problème de la torsion élasto-plastique (2.5),
- le problème de la torsion élasto-plastique en évolution (2.6),
- un problème de frottement évolutif (2.7),
- un problème parabolique à contraintes réparties (2.8),
- un problème hyperbolique à contraintes réparties d'ordre 0 (2.9),
- un problème hyperbolique à contraintes réparties d'ordre 1 (2.10).

2.1 Problème de Dirichlet

Nous allons donner quelques résultats sur la résolution du problème de Dirichlet à une dimension dans le but d'étudier et de comparer un certain nombre de méthodes.

Le problème théorique est

$$\left|\begin{array}{l} \underline{\text{minimiser}}\ J(v) = \displaystyle\int_\Omega |\text{grad } v|^2\, dx - 2\int_\Omega fv\, dx \\ v \in H_o^1(\Omega) \\ \text{avec} \\ \quad \Omega =]0,1[\subset \mathbb{R}, \\ \quad f(x) = 10. \end{array}\right.$$

Ce problème possède la solution exacte

$$u(x) = 5. * x(1.-x)$$

et

$$J(u) = -25./3. = -8.3333$$

Le problème approché s'écrit pour un pas $h = \frac{1}{(N+1)}$, N entier>1,

$$\underline{\text{minimiser}}\ J_h(\vec{v}_h) = \sum_{i=1}^{N} |v_{i+1} - v_i|^2 - 2fh^2 \sum_{i=1}^{N} v_i$$

avec

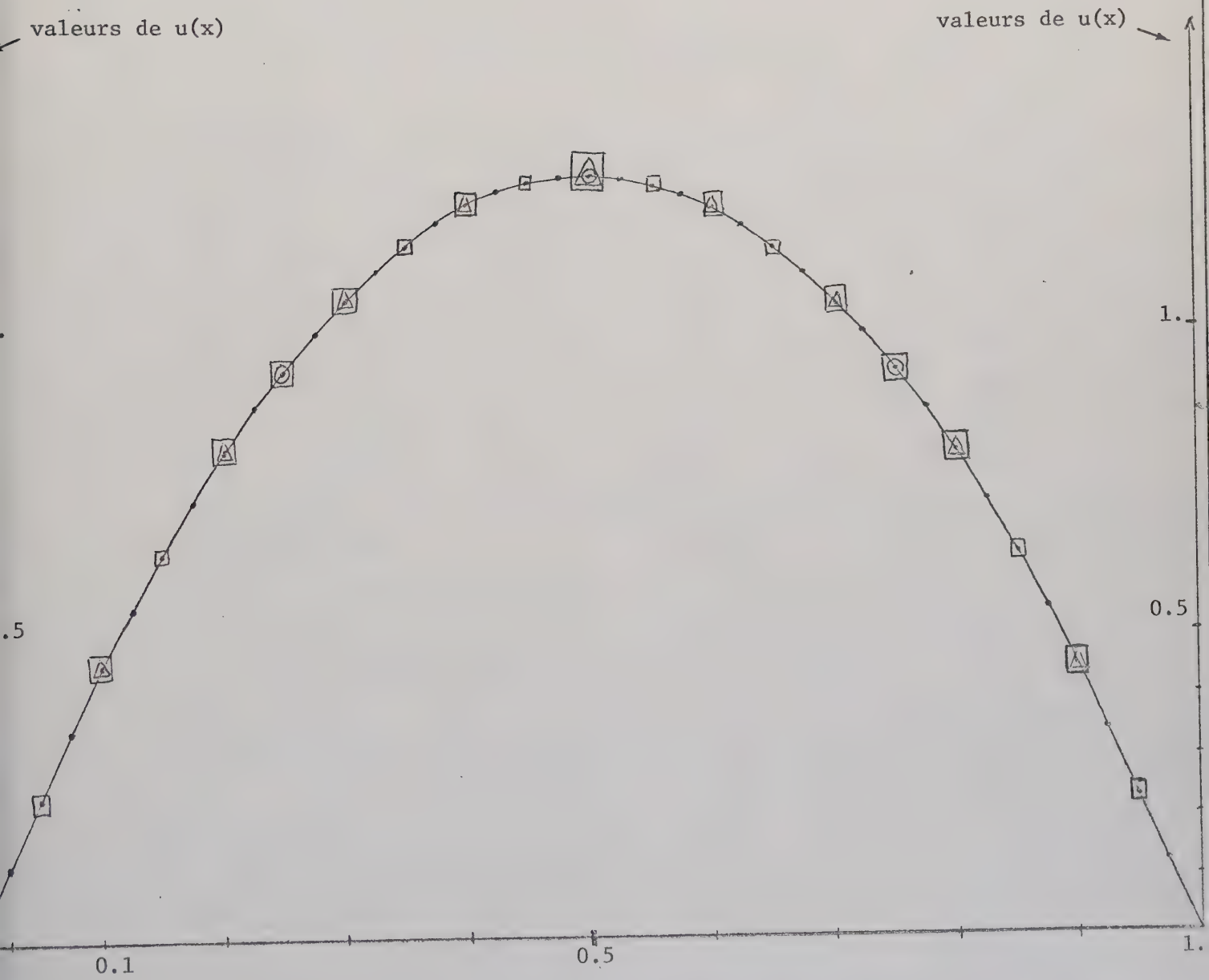
$$\vec{v}_h = (v_1, \ldots, v_N), \qquad v_h = \sum_{i=1}^{N} v_i.\theta^i, \quad v_{N+1} = 0$$

où θ^i est la fonction caractéristique de $](i-\frac{1}{2})h, (i+\frac{1}{2})h[$.

On donne dans le graphique ci-après (Figure 1) la solution exacte et les solutions approchées correspondant à $h = \frac{1}{4}, \frac{1}{10}, \frac{1}{20}, \frac{1}{30}, \frac{1}{40}$.

Figure 1

Comparaison entre la solution exacte du problème de Dirichlet et les solutions discrètes correspondant a h=1/(N+1), pour N=3, 9, 19, 39.



○ valeurs obtenues pour N=3
△ valeurs obtenues pour N=9
□ valeurs obtenues pour N=10
• valeurs obtenues pour N=39
/ solution exacte

Les valeurs calculées numériquement correspondent aux valeurs de la solution exacte a 10^{-8} près.
Calculs effectués sur BULL-GE-265.

Nous allons maintenant étudier le comportement de diverses méthodes appliquées à la résolution du problème de Dirichlet.

a) Valeurs caractéristiques pour la méthode du gradient

(cf. ce chapitre, § 1.7, a). On a, pour $h = \frac{1}{20}$, $N = 19$,

$\lambda_m = 0.044324$, $\lambda_M = 7.94562$,

$\rho_M = 0.25171$, $\boxed{\rho_{opt} = 0.25031}$.

b) Valeurs caractéristiques pour la méthode de relaxation

(cf. ce chapitre, § 1.7, f)).

Le calcul de ω_{opt} par la méthode décrite au § 1.7, f), donne, avec pour critère d'arrêt

$$|\omega_{opt}^{n+1} - \omega_{opt}^{n}| \leq 10^{-5},$$

et pour différentes valeurs de $h = 1/(N+1)$, $N = 5, 10, 20, 30, 40$ les résultats suivants (Tableau 1):

Tableau 1

Valeurs de ω_{opt} pour le problème de Dirichlet sur $\Omega =]0,1[$			
N	ω_{opt} (1)	Nombre d'itérations	Temps (2)
5	1.33334	13	75
10	1.56042	35	131
20	1.73859	7	48
20 (3)	1.74071	87	288
30	1.81693	89	307
40	1.85677	120	420

Commentaires

(1) Le critère d'arrêt est $|\omega_{opt}^{n+1} - \omega_{opt}^{n}| \leq 10^{-5}$.

(2) Le temps est en $1./6^{e}$ de seconde sur BULL-G.E. 265. Ce temps comprend la compilation et le calcul.

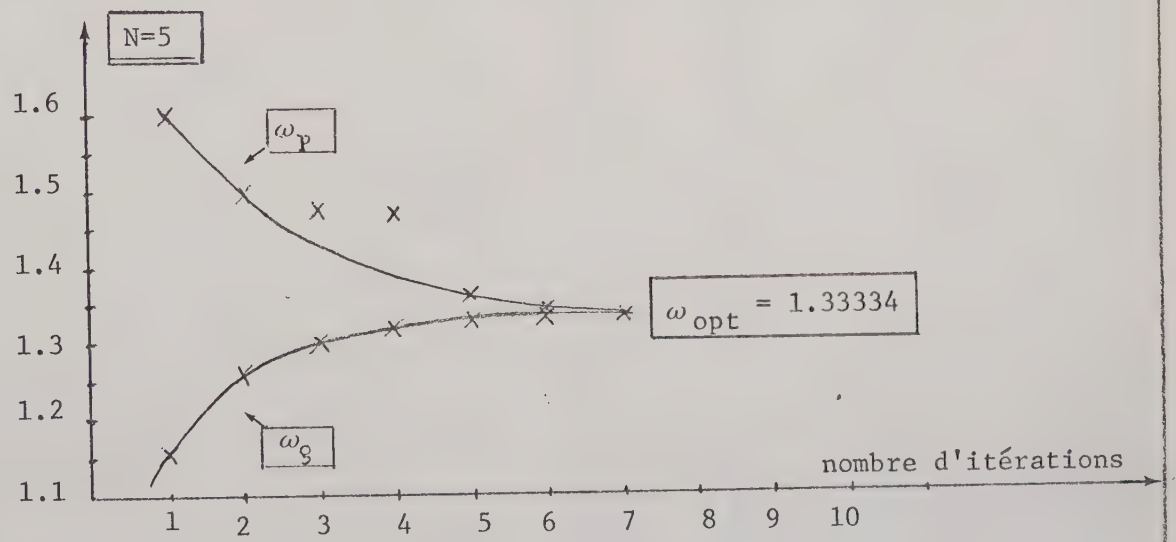
(3) Pour cette ligne le critère d'arrêt a été renforcé à 10^{-6}, le résultat de la ligne précédente n'étant pas satisfaisant.

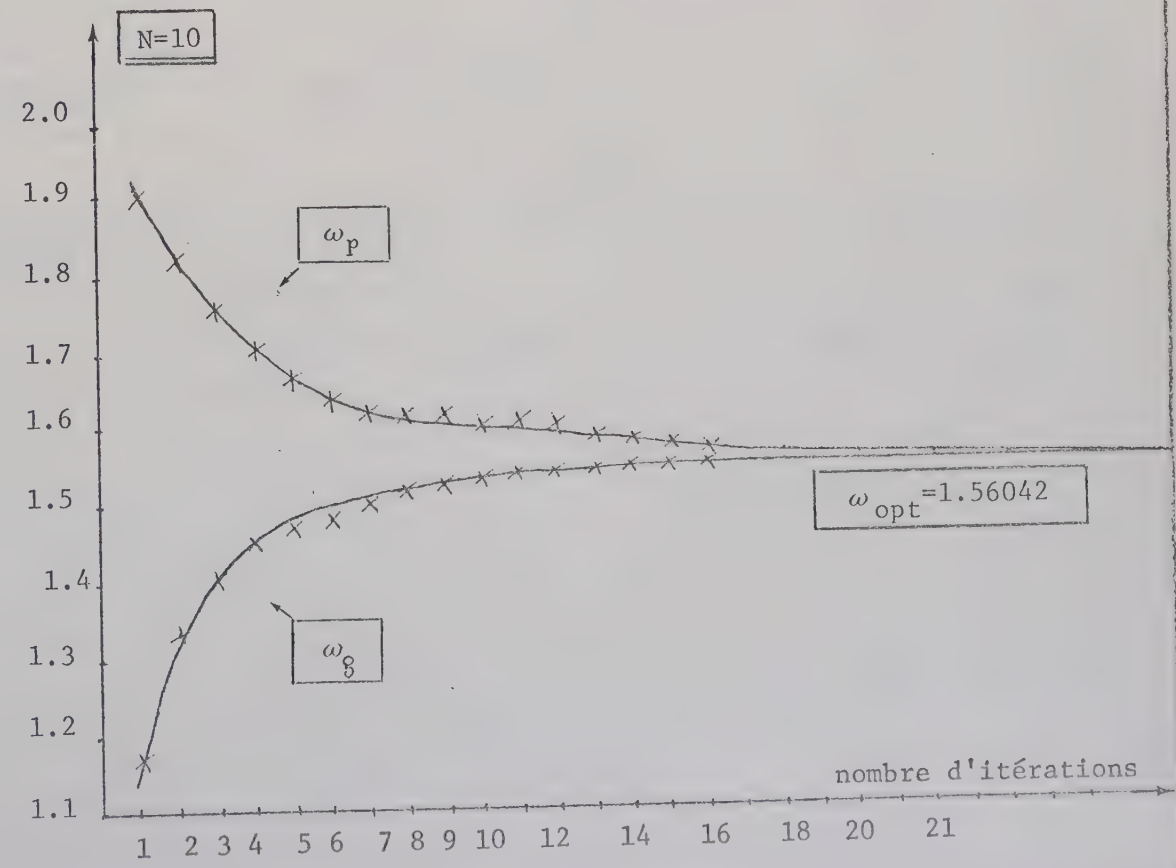
Dans le graphique ci-après (Figure 2) on illustre la convergence des suites ω_p^n et ω_g^n vers la valeur ω_{opt}.

Figure 2

Calcul du paramètre optimal ω_{opt} utilisé dans la méthode de sur-relaxation, pour la résolution du problème de Dirichlet :

$$\begin{cases} -\Delta u = f \text{ , dans } \Omega =]0\text{ , }1[\text{ ,} \\ u(0)=u(1)=0 \text{ .} \end{cases}$$





c) Résultats comparatifs d'ensemble entre les différentes méthodes utilisées

Pour $h = \frac{1}{20}$, N = 19 (nombre de variables) les différentes méthodes utilisées pour la résolution du problème de Dirichlet donnent les résultats consignés dans le tableau ci-après. Le point de départ commun est $v_i = 1.$, $i = 1, \ldots N$.

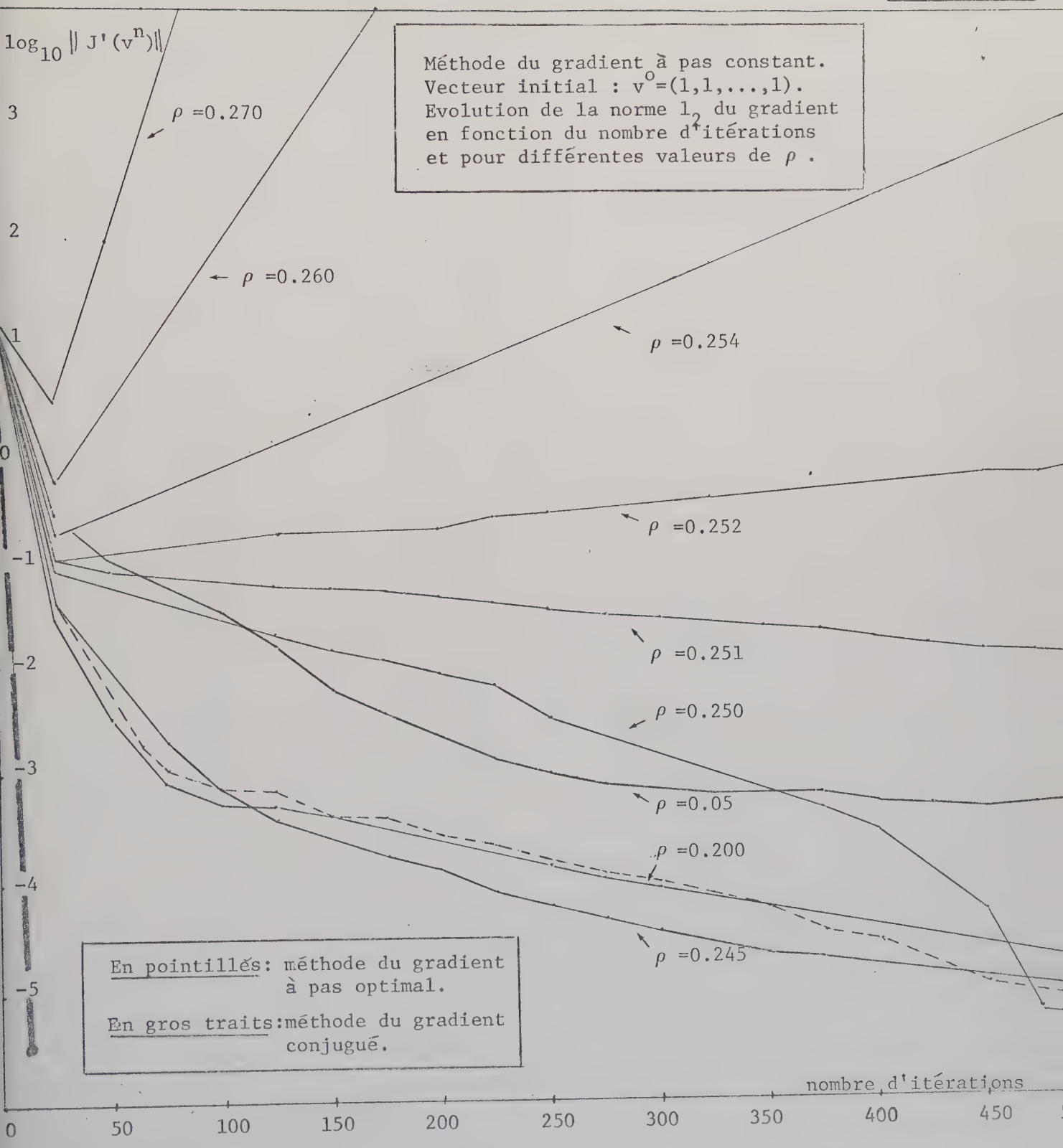
Tableau 2

n = nombre d'itérations T = temps (sec)	n	T	n	T	n	T	n	T
Test d'arrêt	$\|J'(v)\|_2 \leq 10^{-2}$		$\|J'(v)\|_2 \leq 10^{-3}$		$\|J'(v)\|_2 \leq 10^{-4}$		$\|J'(v)\|_2 \leq 10^{-5}$	
Gr.à pas constant = 0.245	50	0.27	150	0.81	275	1.5	500	2.7
Gr. à pas ajusté	30	1.7	40	2.8	150	10.	200	17.
Gr. à pas optimal	40	0.45	100	0.8	240	1.9	380	3.1
Fletcher-Powell	9	0.6	10	0.8	10	0.8	10	0.8
Gauss-Seidel $\omega = 1.73$	18	0.4	26	0.6	35	0.9	47	1.2
Gr. conjugué	9	0.4	10	0.4	10	0.4	10	0.4
Commentaire : Le temps est donné en secondes sur CII 10070.								

d) Importance d'une bonne estimation de ρ dans la méthode du gradient à pas constant

Dans le graphique ci-après (Figure 3), on montre l'importance d'un bon choix du paramètre ρ dans la méthode du gradient à pas constant.

Figure 3


log10 ‖ J'(v^n)‖
Méthode du gradient à pas constant.
Vecteur initial : v^o=(1,1,...,1).
Evolution de la norme l2 du gradient
en fonction du nombre d'itérations
et pour différentes valeurs de ρ .
ρ =0.270
ρ =0.260
ρ =0.254
ρ =0.252
ρ =0.251
ρ =0.250
ρ =0.05
ρ =0.200
ρ =0.245
En pointillés: méthode du gradient à pas optimal.
En gros traits: méthode du gradient conjugué.
nombre d'itérations
3
2
1
0
-1
-2
-3
-4
-5
0
50
100
150
200
250
300
350
400
450

e) Variation du "paramètre" dans la méthode du gradient à pas optimal

On donne dans le Tableau 3 le détail des itérations dans la méthode du gradient à pas optimal. Le point de départ est $v^o = (1, 1, \ldots, 1)$.

On notera que pour les valeurs de la fonction J(v) éloignées du minimum, la valeur de ρ^{n+1} est peu différente de la valeur du pas fixe "optimal" $\rho_{opt} = 0,250$.

Tableau 3

Nombre d'itérations n	$J_h(v^n)$	Norme l_2 de $J_h'(v^n)$	Norme l_2 de $(u-v^n)$	Temps d'exécution sur CII-10070	ρ_n
1	+2,3	1,5	1,1	0,008 s	0,245
5	-0,21	0,65	0,82		0,245
10	-7,65	0,14	0,38	0,08 s	0,244
20	-8,24	0,042	0,13		0,240
40	-8,311	5×10^{-3}	0,04		0,239
60	-8,3120	$1,2\times10^{-3}$	0,03		0,248
80	-8,31221	$8,2\times10^{-4}$	0,024		0,250
100	-8,31233	7×10^{-4}	0,018	0,8 s	0,252
120	-8,31239	4×10^{-4}	0,014		0,258
140	-8,31243	$3,8\times10^{-4}$	0,011		0,261
160	-8,31246	3×10^{-4}	0,0088		0,274
180	-8,31246	$2,3\times10^{-4}$	0,0080		0,278
200	-8,31247	$1,8\times10^{-4}$	0,0054	1,6 s	0,292
220	-8,31248	$1,3\times10^{-4}$	0,0042		0,324
240	-8,31248	10^{-4}	0,0033		0,343
260	-8,31248	$7,3\times10^{-5}$	0,0026		0,420
280	-8,31248	$5,4\times10^{-5}$	0,0020		0,446
300	-8,31248	$3,5\times10^{-5}$	0,0015	2,4 s	0,767
320	-8,31248	$2,6\times10^{-5}$	1×10^{-3}		0,490
340	-8,31249	$1,6\times10^{-5}$	6×10^{-4}		0,593
360	-8,31249	$1,2\times10^{-5}$	5×10^{-4}		0,494
380	-8,31249	9×10^{-6}	4×10^{-4}		0,624
400	-8,31249	$8,1\times10^{-6}$	1×10^{-4}	3,2 s	0,306

f) Importance d'un bon choix de ω dans la méthode de relaxation

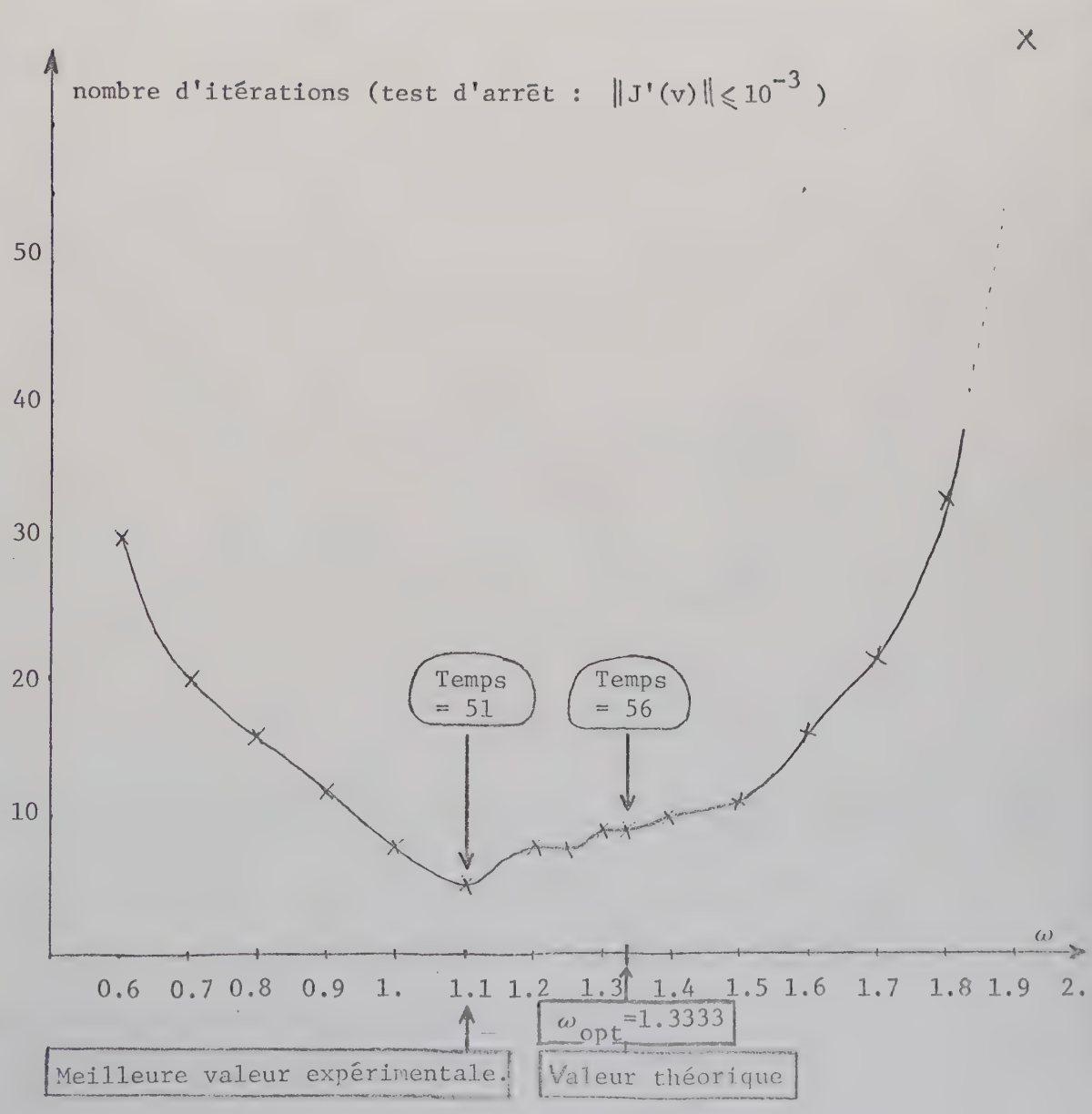
Dans les deux graphiques ci-après (Figures 4 et 5) on donne pour N = 5 et N = 10, le nombre d'itérations nécessaires pour obtenir

$$\| J_h'(v) \|_2 \leqslant 10^{-3}$$

en fonction de plusieurs choix de ω .

Figure 4

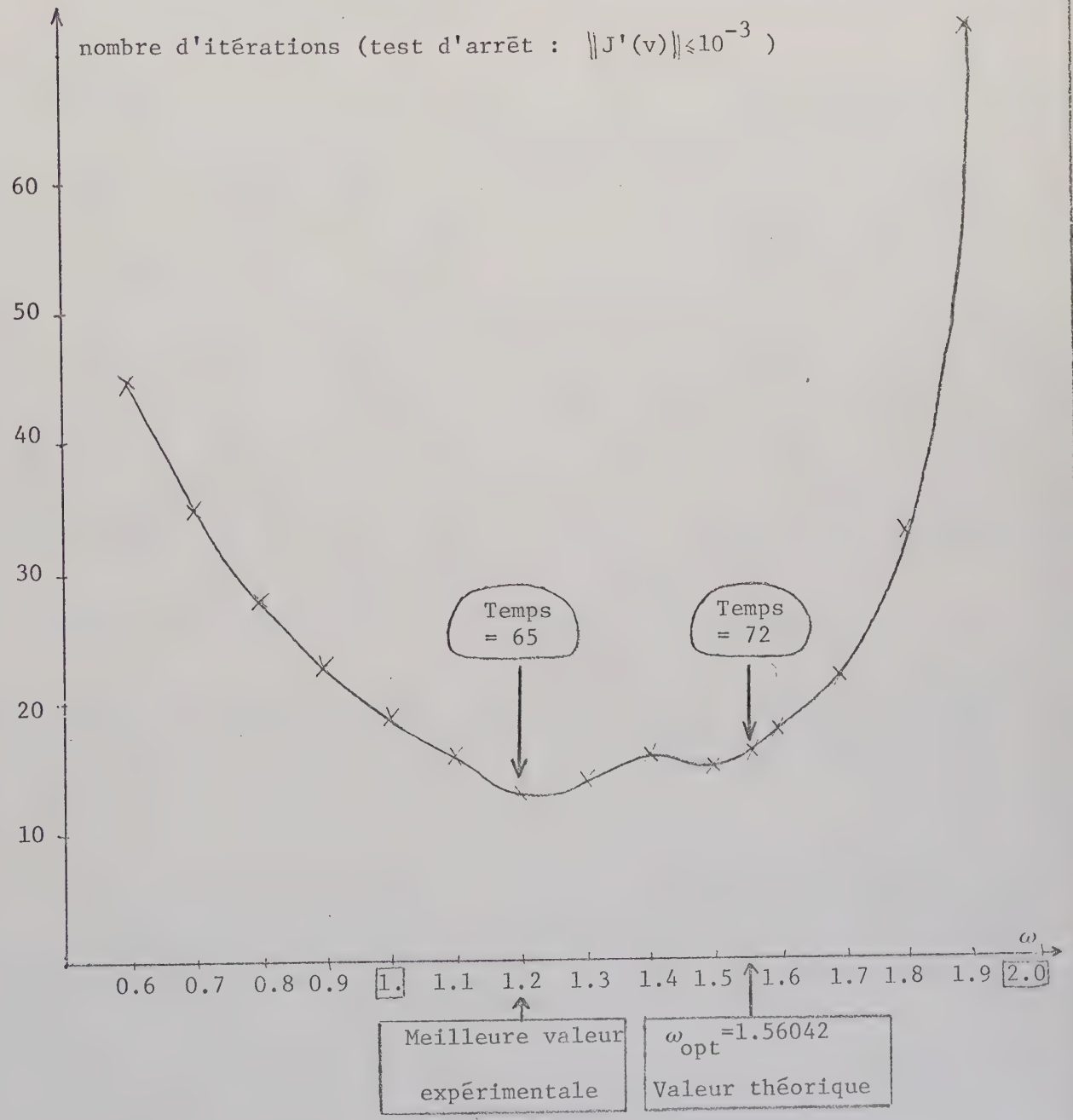
Variation de la rapidité de la méthode de Gauss-Seidel (appliquée au problème de Dirichlet) en fonction de ω. Les résultats sont donnés pour N=5 (5 variables) ; vecteur initial : $v_i = 1.$, i=1,...,5 .



Programmé en Time-sharing sur BULL-GE-265.

Figure 5

Variation de la rapidité de la méthode de Gauss-Seidel (appliquée au problème de Dirichlet) en fonction de ω. Les résultats sont donnés pour N=10 (10 variables) ; vecteur initial : $v_i=1.$, $i=1,\ldots,10$.



Programmé en Time-sharing sur BULL-GE-265 .

g) Comparaison entre les méthodes de Gauss-Seidel et du gradient conjugué.

Dans le Tableau 4 ci-après on donne pour $h = 1./(N+1)$ avec $N = 5, 10, 20, 30, 40$ des éléments de comparaison pour les méthodes de Gauss-Seidel (G.S.) et du gradient conjugé (G.CO.). Le point de départ commun est $v_i = 1.$, $i = 1, \ldots, N$ et le test d'arrêt est

$$\|J'_h(v)\|_2 \leqslant 10^{-3}.$$

La méthode de Gauss-Seidel a été programmée avec les valeurs données dans le Tableau 1.

Tableau 4

	Temps [1]		Nombre d'itérations		Valeur de J(v)	
N	G.S.	G.CO.	G.S.	G.CO.	G.S.	G.CO.
5	56	65 [3]	9	3	-4.050926	-4,050926
10	72	67	16	5	-4.132228	-4.132231
20	109	83	29	10	-4.157198	-4.157219
30	152	103	41	14	-4.162285	-4.162331
40	207	126	53	19	-4.164116	-4.164188
40 [2]	254	126	81	19	-4.164188	-4.164188

Commentaires : [1] Le temps est en $1./6^e$ de seconde sur BULL-GE-265 et comprend la compilation et le calcul.

[2] Cette ligne correspond à un test d'arrêt renforcé à $\|J'_h(v)\|_2 \leqslant 10^{-4}$.

[3] Pour N = 5, le temps supérieur pris par la méthode du gradient conjugué s'explique en partie par le fait que la programmation du gradient conjugué est un peu plus longue que celle de la méthode de Gauss-Seidel.

Le Tableau 5 qui suit correspond aux mêmes données que celles du Tableau 4, la seule différence réside dans le point de départ que l'on a pris ainsi $v_i = i - 1$, $i = 1, \ldots, N$.

Tableau 5

N	Temps		Nombre d'itérations		Valeur de J(v)	
	G.S.	G.CO.	G.S.	G.CO.	G.S.	G.CO.
5	56	70	11	5	-4.050924	-4.050926
10	64	75	21	10	-4.132220	-4.132231
20	94	100	41	20	-4.157177	-4.157219
30	174	135	62	30	-4.162266	-4.162331
40	262	181	82	40	-4.163935	-4.164188

Dans les graphiques ci-après (Figures 6 et 7) on a représenté la variation de $J(v^n)$ en fonction de n = 1,2, ... obtenu par les méthodes de Gauss-Seidel et du gradient conjugué.

Figure 6

Valeurs de $J(v^n)$ en fonction de n = 1,2,... pour les méthodes de Gauss-Seidel et du gradient conjugué.
Résultats pour le problème de Dirichlet avec N = 10 .
Vecteur initial : $v_i = 1$, i = 1,...,10 .
Test d'arrêt : $\|J'(v^n)\|_2 \leq 10^{-3}$.

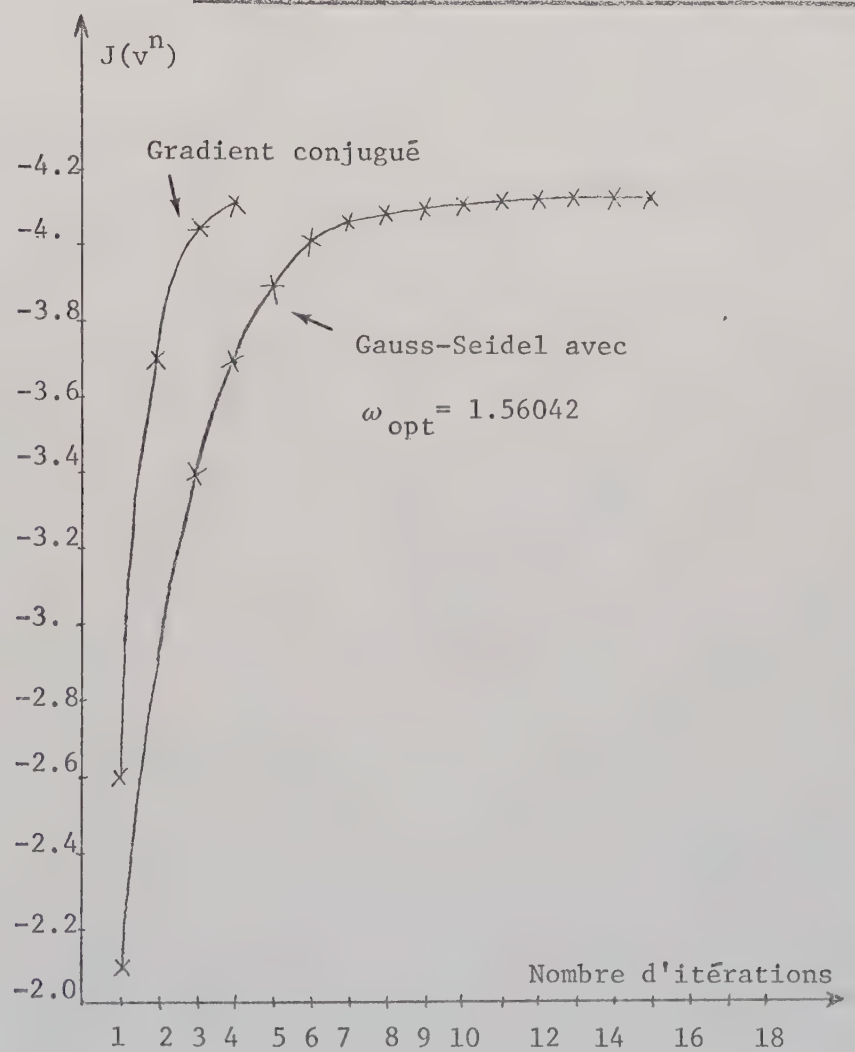
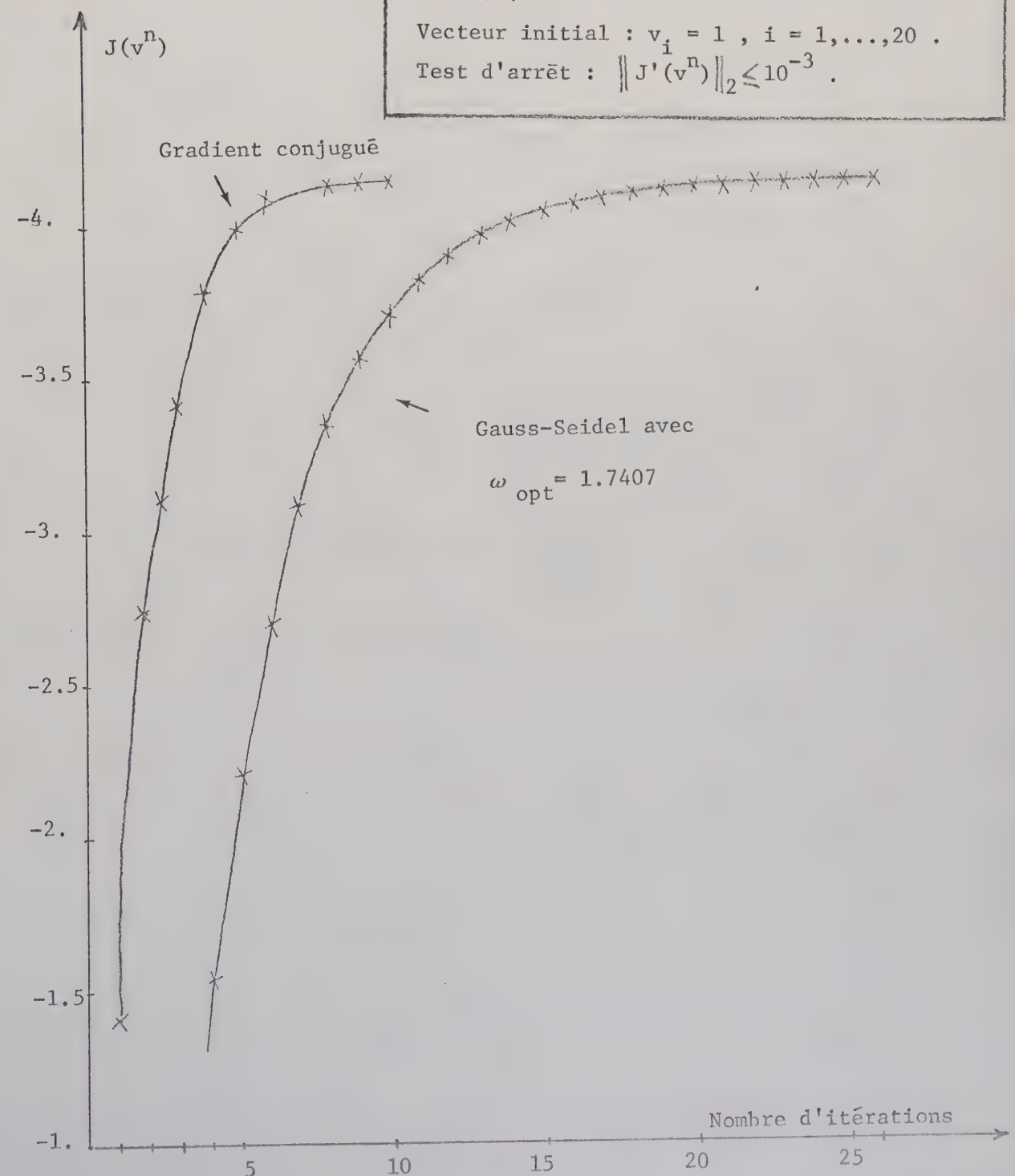


Figure 7

Valeurs de $J(v^n)$ en fonction de n = 1,2,... pour les méthodes de Gauss-Seidel et du gradient conjugué.
Résultats pour le problème de Dirichlet avec N = 20 .
Vecteur initial : $v_i = 1$, i = 1,...,20 .
Test d'arrêt : $\| J'(v^n) \|_2 \leq 10^{-3}$.



h) Conclusions de 2.1

Les enseignements que nous pouvons tirer sur l'intérêt relatif des diverses méthodes ne peuvent en aucun cas prétendre à la généralité. Il ressort cependant de ce 1er paragraphe les conclusions suivantes :

1 - Méthode du gradient à pas fixe
- programmation très facile ,
- convergence extrêmement sensible au choix du pas fixe,
- difficulté de la détermination du pas optimal ,
- convergence bonne pour $\rho = \rho_{opt.fixe}$, cependant assez lente dans la zone du minimum.

2 - Méthode du gradient à pas ajusté
- programmation assez simple ,
- (évite la recherche d'un pas "optimal"),
- convergence en moins d'itérations que la méthode précédente mais
- plus coûteuse en temps.

3 - Méthode du gradient à pas optimal
- programmation assez simple ,
- valable pour les problèmes où le minimum dans une direction quelconque peut être déterminé explicitement ,
- convergence tout à fait identique que celle de la 1ère méthode mais
- sensiblement équivalente en temps à la méthode précédente.

4 - Méthode de Fletcher-Powell
- programmation extrêmement délicate ,
- inutilisable pour de gros problèmes lorsque le calculateur a peu de mémoires,
- convergence extrêmement précise ,
- peu coûteuse en temps puisque 5 fois plus rapide que la "meilleure" méthode de gradient.

5 - Méthode de Gauss-Seidel
- programmation très facile ,
- méthode nécessitant très peu de mémoires ,
- difficulté et coût supplémentaire pour le calcul du paramètre ω_{opt} ; le temps de calcul de ω_{opt} (relatif au problème de Dirichlet) par la méthode que nous avons utilisée est sensiblement le même que le temps de ré-

solution du problème lui-même. Cependant une estimation assez large et par valeur inférieure de $\omega_{opt.théorique}$ ne perturbe guère le temps de résolution du problème de Dirichlet,

- convergence très bonne (mais moins précise et moins rapide que la méthode précédente).

6 - Méthode du gradient conjugué

- programmation assez délicate ,
- demande peu de mémoires, mais plus que la méthode précédente (de 2 à 4 fois plus suivant la manière de programmer),
- "pas de paramètre à calculer",
- convergence extrêmement rapide, très précise et demandant peu d'itérations (en général le nombre d'itérations est $< N$),
- sans conteste la meilleure des méthodes étudiées ici. ■

2.2 Problème de Neumann

Le problème théorique considéré ici est

$$\left|\begin{array}{l} \underline{\text{minimiser}}\ J(v) = \displaystyle\int_\Omega u^2\,dx + \int_\Omega |\text{grad } u|^2\,dx - 2\int_\Omega fv\,dx \\ v \in H^1(\Omega) \\ \Omega =]0,1[\\ f(x) = u - \Delta u \text{ avec} \\ u(x) = 10.*x^2*(1.-x^2). \end{array}\right. \blacksquare$$

Le problème approché pour un pas de discrétisation de $h = 1./(N-1)$, N entier > 1 s'écrit

$$\left|\begin{array}{l} \underline{\text{minimiser}}\ J_h(v_h) = \left(\dfrac{|v_1|^2}{2} + \dfrac{|v_N|^2}{2} + \displaystyle\sum_{i=2}^{N-1}|v_i|^2 + \sum_{i=2}^{N}\frac{|v_i - v_{i-1}|^2}{h^2}\right. \\ \qquad \left. - f(o)v_1 - f(N)v_N - 2\displaystyle\sum_{i=2}^{N-1} f(i)v_i\right) * h \end{array}\right.$$

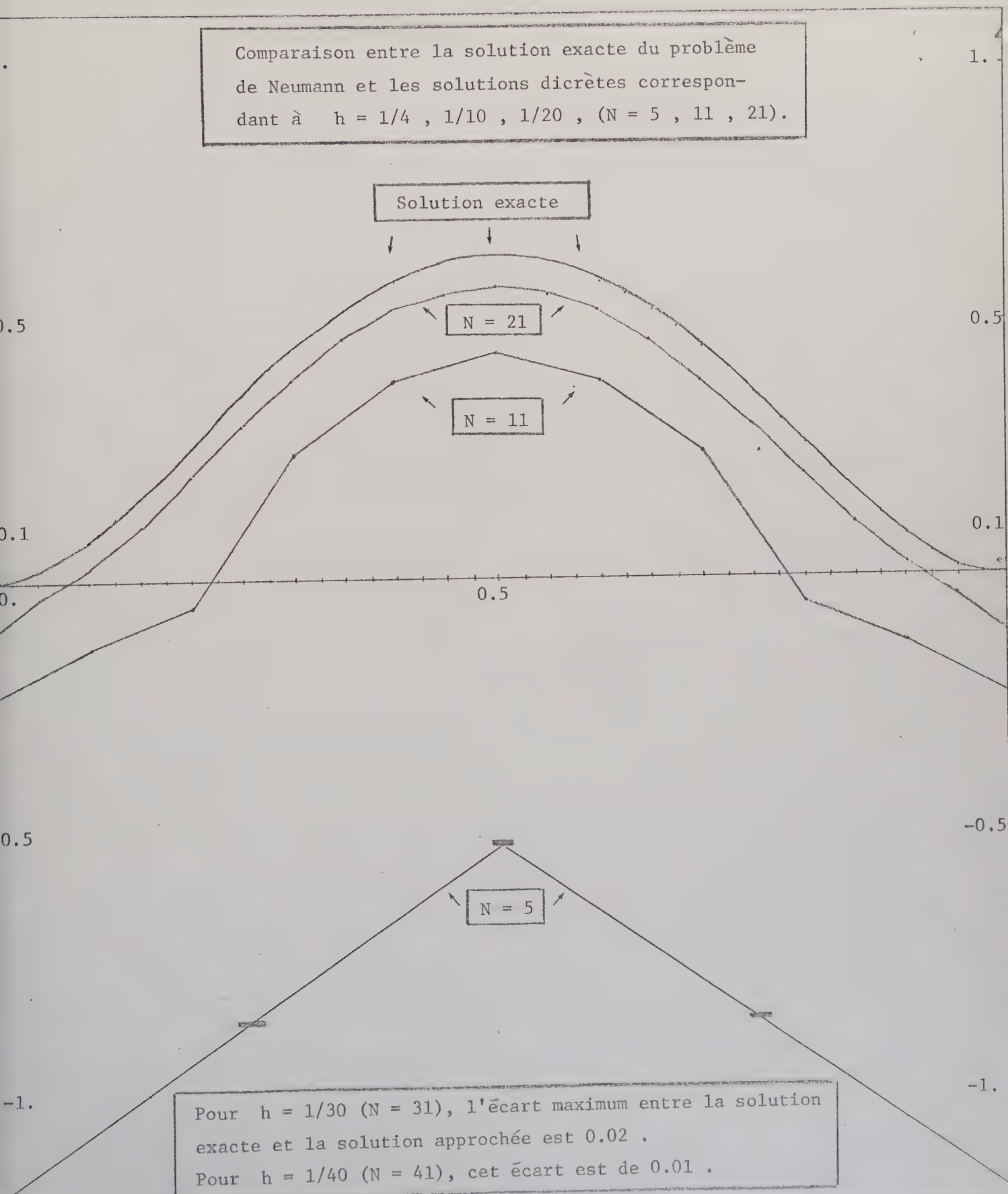
La solution approchée est

$$u_h = \sum_{i=2}^{N-1} u_i\theta^i + u_1\theta^1 + u_N\theta^N$$

où θ^i, $i = 2, \ldots, N-1$ est la fonction caractéristique de $](i - \frac{1}{2})h, (i + \frac{1}{2})h[$ et θ^1 celle de $]0, \frac{1}{2}h[$, θ^N celle de $](N - \frac{1}{2})h, Nh[$. ■

On donne dans le graphique ci-après (Figure 8) la solution exacte et les solutions approchées correspondant à $h = \frac{1}{4}$, $\frac{1}{10}$, $\frac{1}{20}$, $\frac{1}{30}$, $\frac{1}{40}$.

Figure 8



a) Calcul du paramètre optimal pour la méthode de Gauss-Seidel

Le calcul de ω_{opt} par la méthode décrite au § 1.7, f) donne, avec pour critère d'arrêt

$$\left|\omega_{opt}^{n+1} - \omega_{opt}^{n}\right| \leqslant 10^{-5} \quad \text{et} \quad n \geqslant 2.* N$$

les résultats suivants :

Tableau 6

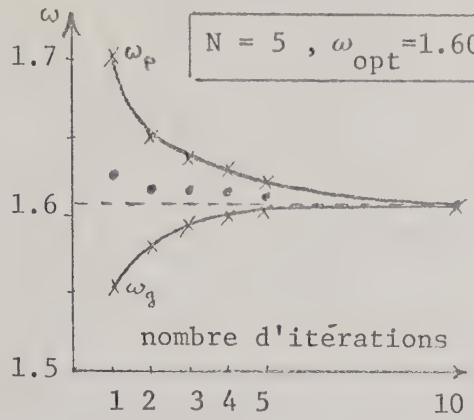
Valeurs de ω_{opt} pour le problème de Neumann sur $\Omega =]0,1[$			
N	ω_{opt}	Nombre d'itérations	Temps
5	1.60732	14	36
10	1.80090	37	55
20	1.90030	41	78
30	1.93342	61	84
40	1.95002	81	130

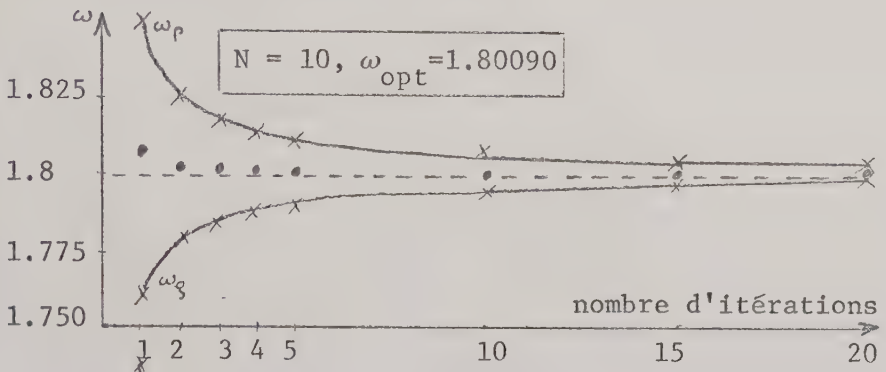
La convergence des suites ω_p^n et ω_β^n sont illustrées sur le graphique ci-après (Figure 9):

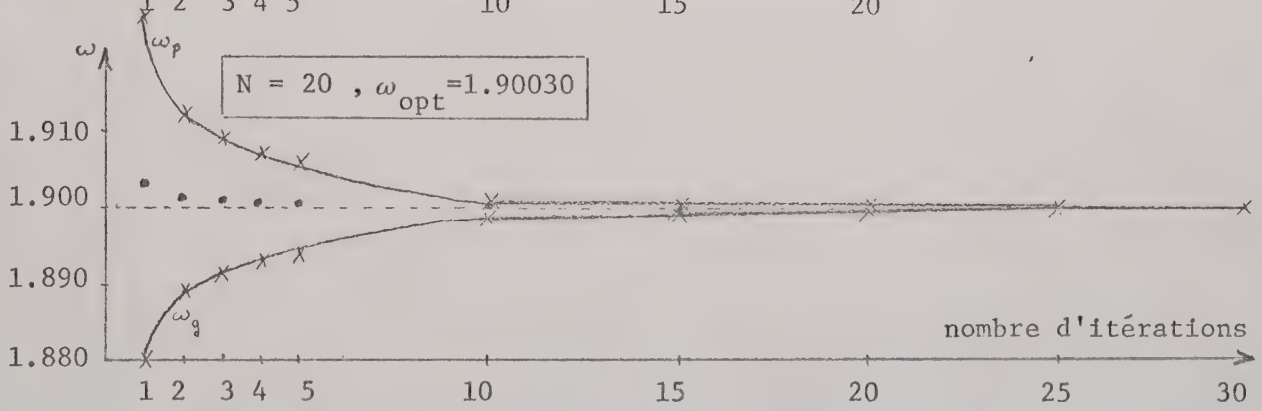
Figure 9

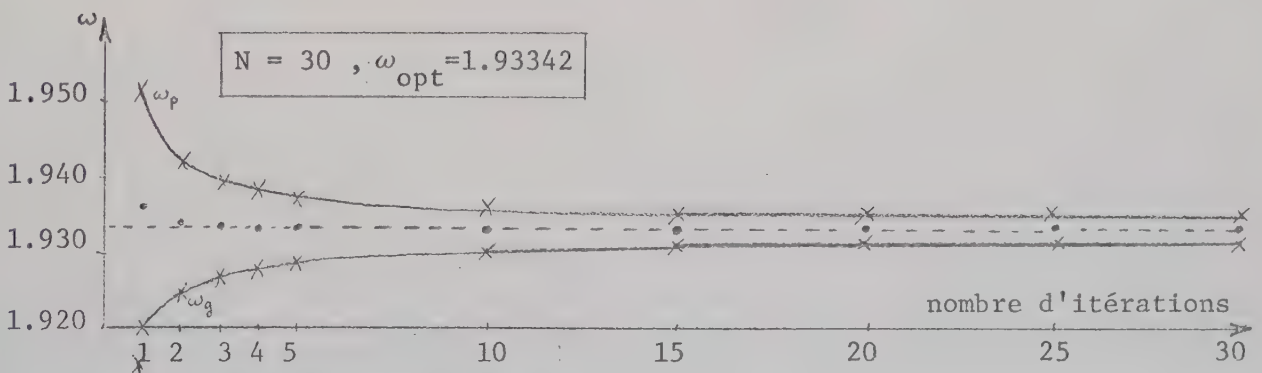
Calcul du paramètre optimal ω_{opt} utilisé dans la méthode de sur-relaxation pour la résolution du problème de Neumann :

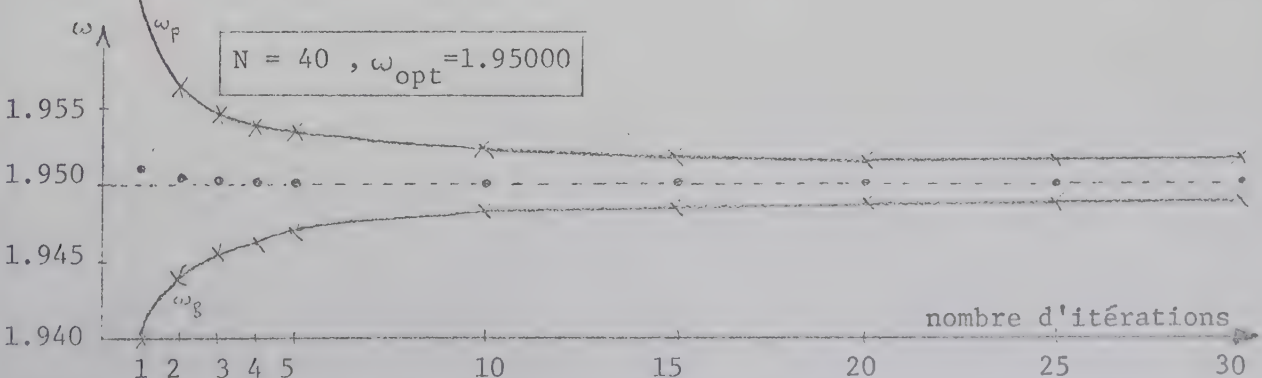
$$\begin{cases} u - \Delta u = f, \text{ dans } \Omega =]0\ ,\ 1[\ , \\ u(0) = \quad u(1) = 0\ . \end{cases}$$











b) Importance d'un bon choix de ω dans la méthode de relaxation

Dans les trois graphiques ci-après (Figures 10, 11, 12) on donne pour N = 5, 10, 20, 30, 40, le nombre d'itérations nécessaires pour obtenir

$$\|J'(v)\|_2 \leq 10^{-4} .* h$$

dans la résolution du problème de Neumann, en fonction de plusieurs choix de ω .

Figure 10

Variation de la rapidité de la méthode de Gauss-Seidel (appliquée au problème de Neumann) en fonction de ω.
Les résultats sont donnés pour N = 5 (5 variables) ;
vecteur initial : $v_i = 1.$, $i = 1, \ldots, 5.$
Test d'arrêt : $\|J'(v)\|_2 \leq 10^{-4} . h$.

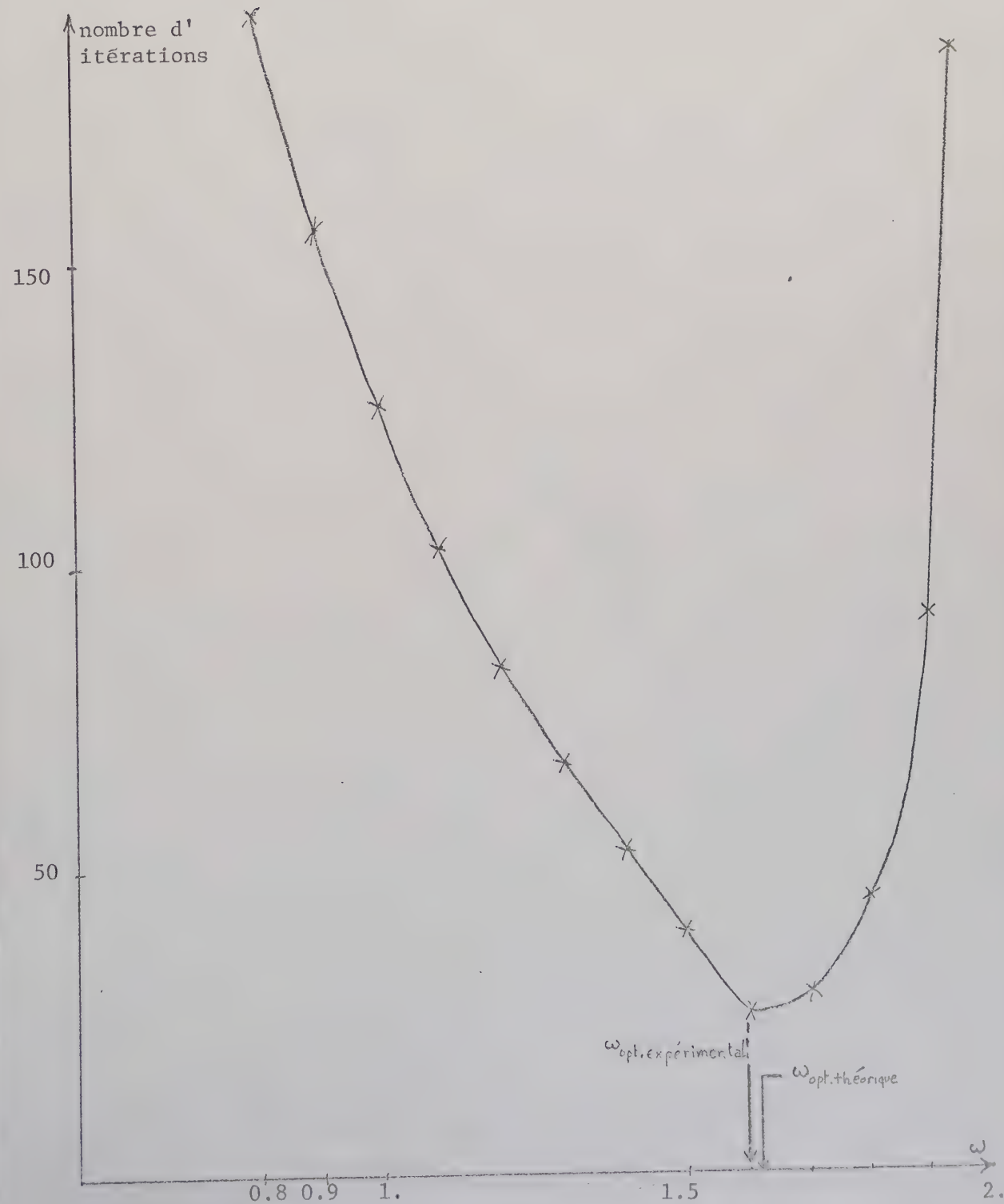


Figure 11

Variation de la rapidité de la méthode de Gauss-Seidel (appliquée au problème de Neumann) en fonction de ω .
Les résultats sont donnés pour N = 10 (10 variables);
vecteur initial : $v_i = 1.$, $i = 1, \ldots, 10.$
Test d'arrêt : $\| J'(v) \|_2 \leq 10^{-4} . h$.

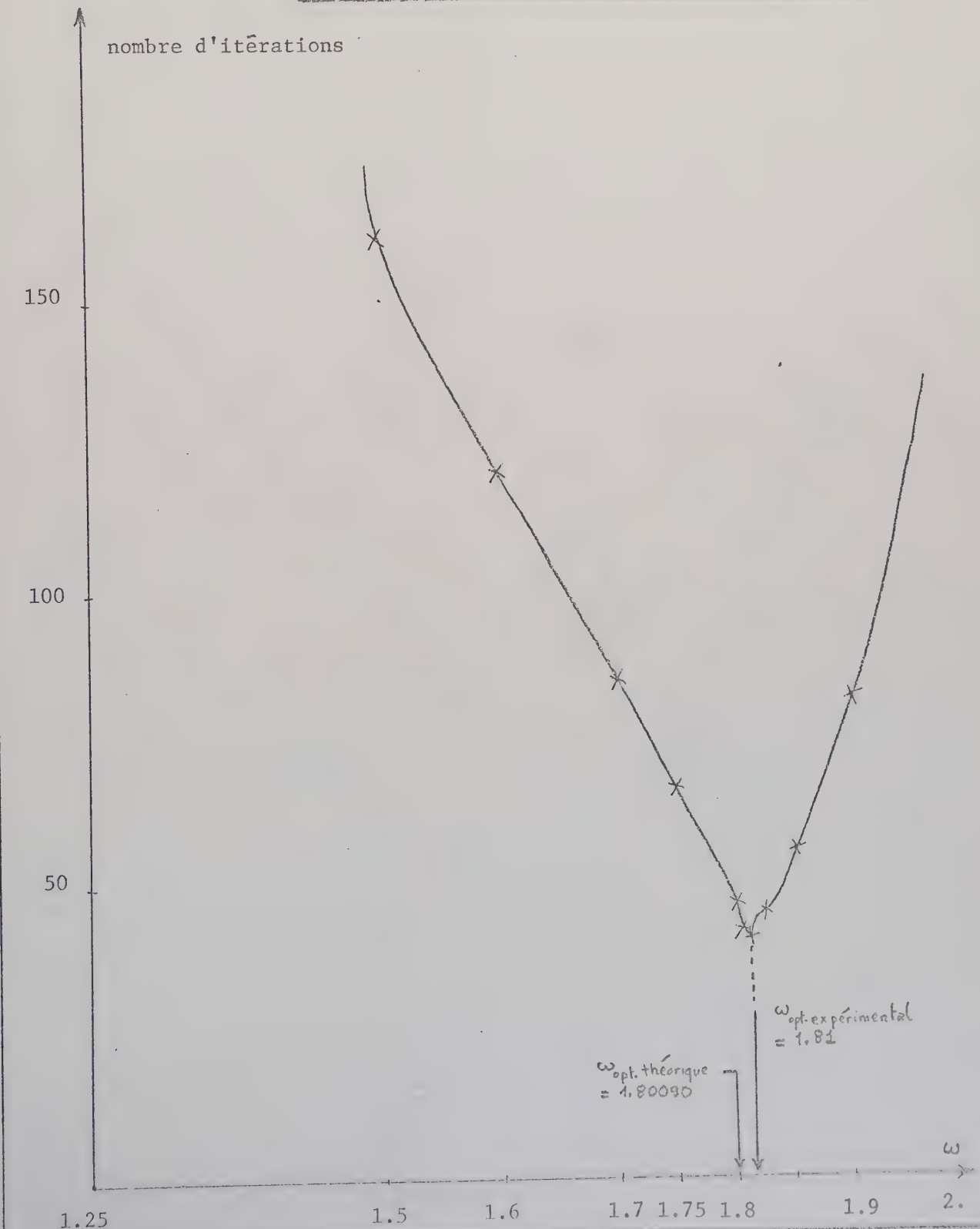
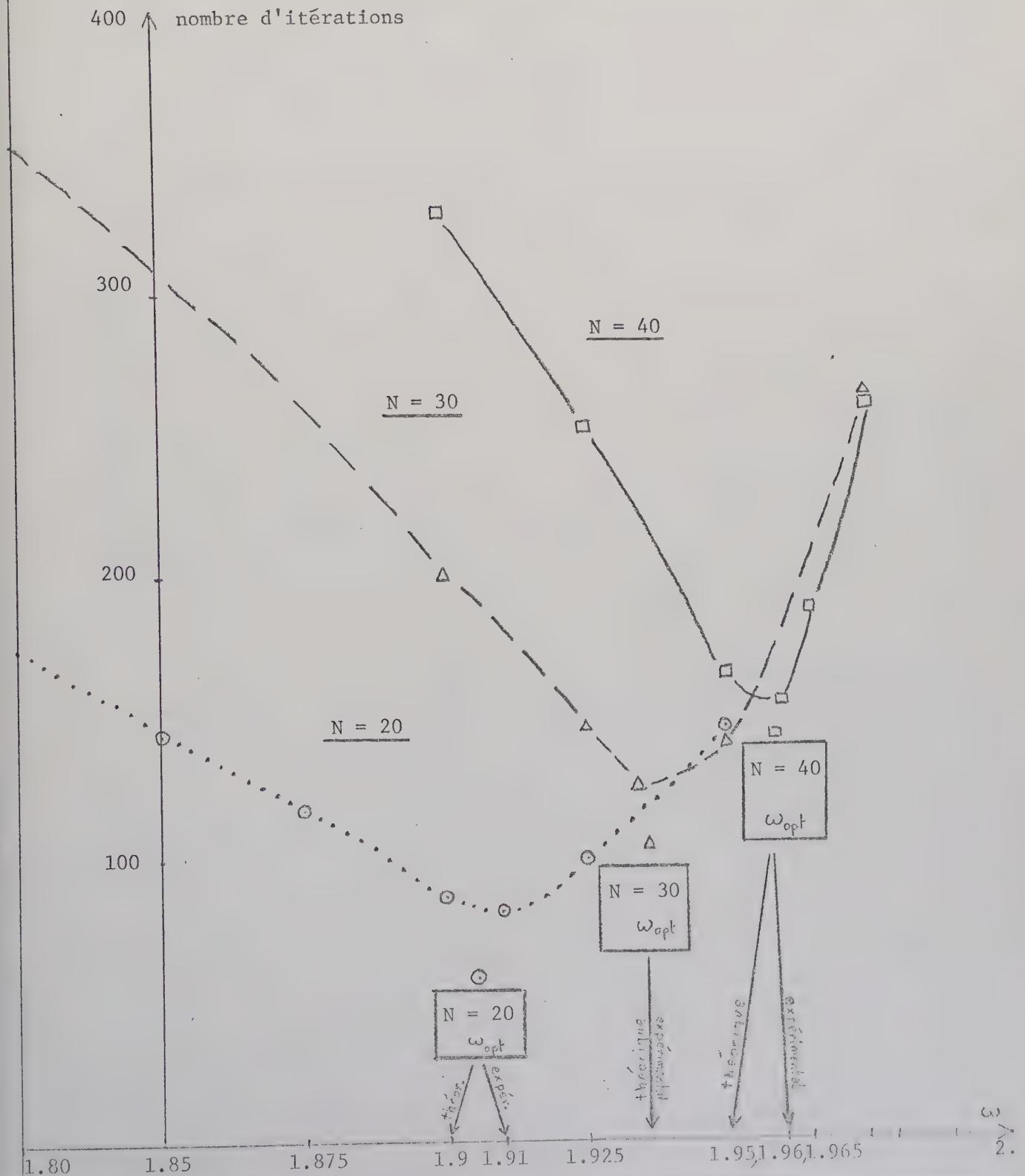


Figure 12

Variation de la rapidité de la méthode de Gauss-Seidel (appliquée au problème de Neumann) en fonction de ω .
Les résultats sont donnés pour N = 20 , N = 30 , N =40;
vecteur initial : $v_i = 1.$, $i = 1, \ldots, N$.
Test d'arrêt : $\| J'(v) \|_2 \leqslant 10^{-4} . h$.



c) Comparaison entre les méthodes de Gauss-Seidel et du gradient conjugué

Dans le Tableau 7 ci-après on donne pour h = 1./(N+1) avec N = 5, 10, 20, 30, 40, des éléments de comparaison pour les méthodes de Gauss-Seidel (G.S.) et du gradient conjugué (G.CO.). Le point de départ commum est $v_i = 1.$, i = 1, ..., N et le test d'arrêt est

$$\|J'(v)\|_2 \leq 10^{-4} \cdot h .$$

La méthode de Gauss-Seidel a été programmée avec les valeurs ω_{opt} données par le Tableau 6.

Tableau 7

	Temps[1]		Nombre d'itérations		Valeur de J'(v) . h		Valeur de J(v) . h
N	G.S.	G.CO.	G.S.	G.CO.	G.S.	G.CO.	G.S. G.CO. (valeurs égales
5	44	54	24	3	9.10^{-5}	2.10^{-8}	-1.15347
10	52	54	47	5	9.10^{-5}	1.10^{-6}	-0.24783
20	93	57	86	10	9.10^{-5}	2.10^{-6}	-0.10980
30	145	58	126	15	1.10^{-4}	1.10^{-5}	-0.07146
40	224	79	165	20	1.10^{-4}	2.10^{-5}	-0.05303

Commentaire : [1] Le temps est donné en $1/6^e$ de seconde sur BULL-GE-265 ; le temps comprenant la compilation et le calcul.

Dans les graphiques qui suivent (Figures 13 et 14) on illustre la convergence des méthodes de Gauss-Seidel et du gradient conjugué.

Figure 13

Variation de la norme du gradient pour les méthodes de Gauss-Seidel (avec $\omega = \omega_{opt.th.}$) et du gradient conjugué, appliquées a la résolution du problème de Neumann : N = 20 , h = 1/19 ; le vecteur initial est $v_i = 1.$, $i = 1, \ldots, N$.

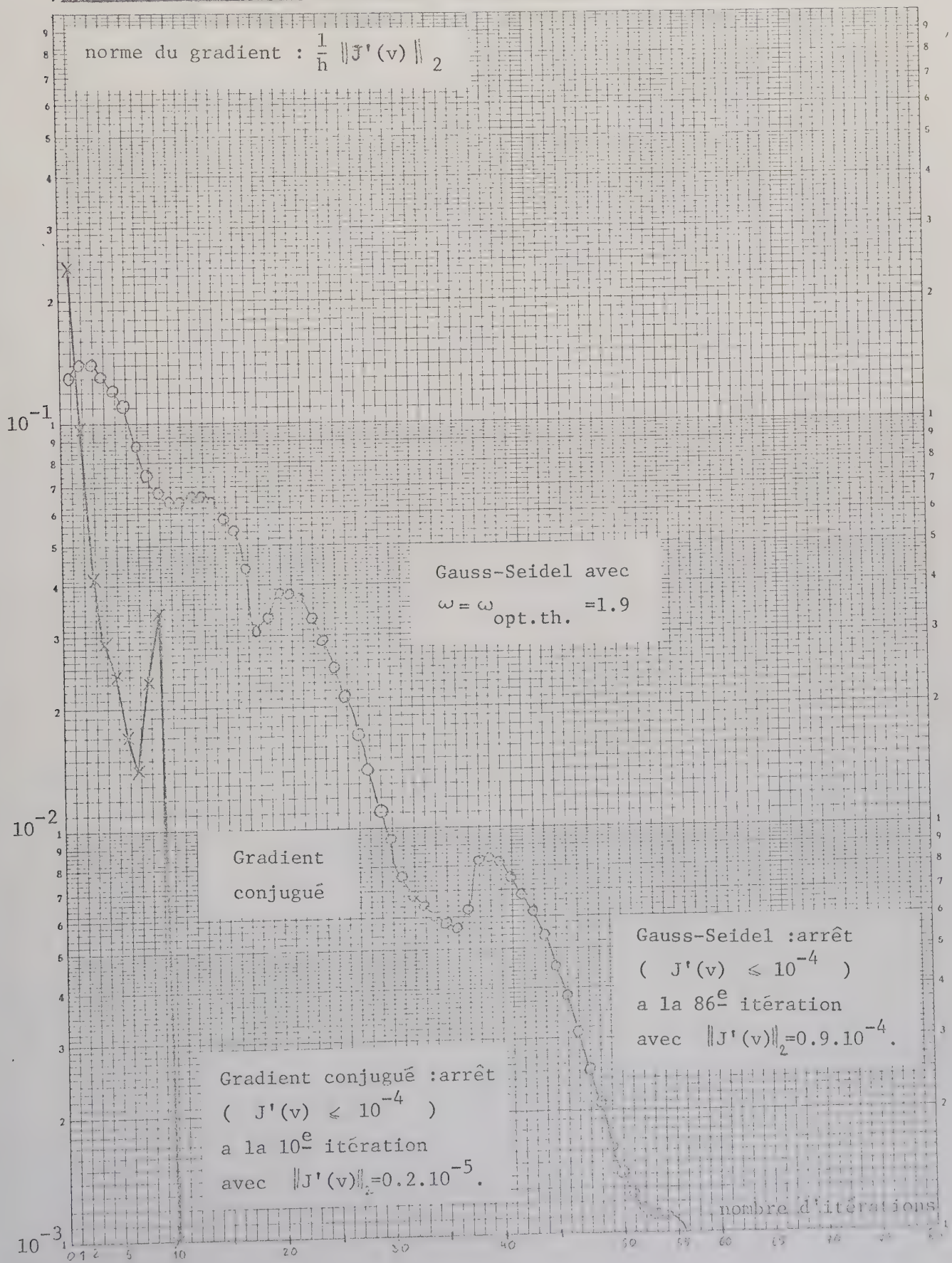
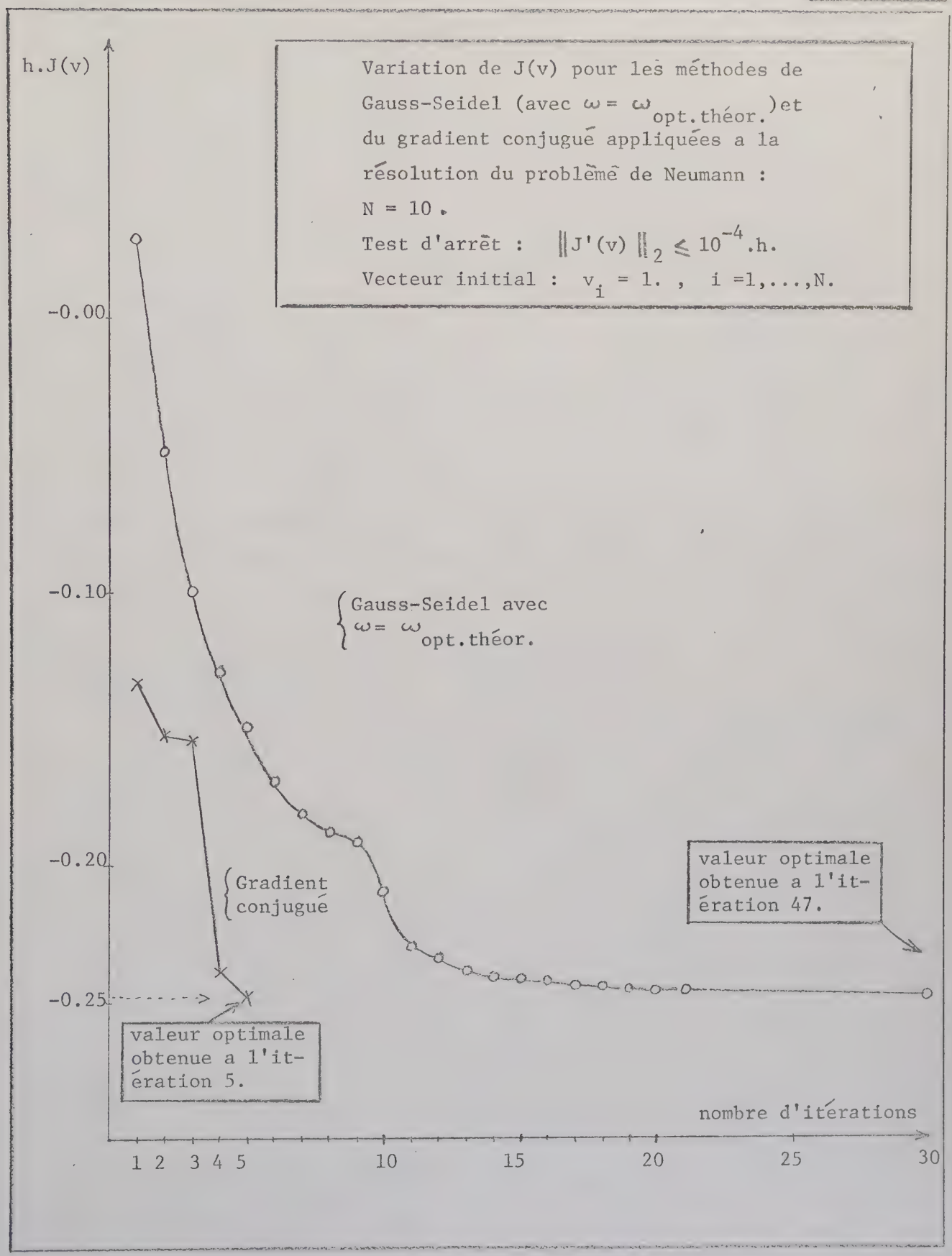


Figure 14


h.J(v)
Variation de J(v) pour les méthodes de Gauss-Seidel (avec ω = ω opt.théor.) et du gradient conjugué appliquées a la résolution du problème de Neumann :
N = 10.
Test d'arrêt : ‖J'(v)‖2 ≤ 10^-4.h.
Vecteur initial : v_i = 1. , i =1,...,N.
-0.00
-0.10
-0.20
-0.25
Gauss-Seidel avec ω = ω opt.théor.
Gradient conjugué
valeur optimale obtenue a 1'itération 47.
valeur optimale obtenue a 1'itération 5.
nombre d'itérations
1 2 3 4 5 10 15 20 25 30

d) Conclusions

Dans ce paragraphe nous avons voulu savoir si les résultats obtenus pour le problème de Dirichlet résolu par la méthode de relaxation et par la méthode du gradient conjugué étaient encore valables pour le problème de Neumann. Les résultats obtenus le confirment.

Nous pouvons maintenant conclure en ce qui concerne la méthode de Gauss-Seidel appliquée aux problèmes de Dirichlet et de Neumann :

- le calcul d'un $\omega_{opt.théor.}$ précis pour le problème de Dirichlet ne semble pas nécessaire, et fort heureusement d'ailleurs car le calcul de $\omega_{opt.théor.}$ est long.
- ce calcul est fondamental pour le problème de Neumann et très heureusement il s'évalue extrêmement vite pour ce problème (pour des résultats dans ce sens cf. BOURGAT-TREMOLIERES [1]).

Quant à la méthode du gradient conjugué elle est dans tous les cas (sauf si le nombre de variables est < 10) plus rapide que la méthode de Gauss-Seidel. ■

2.3 Un problème de climatisation stationnaire

On se propose de

$$\text{minimiser } J(v) = \frac{1}{2}\int_\Omega (v^2 + \text{grad } v^2)dx - \int_\Omega fv\,dx + \int_\Gamma \psi(v)d\Gamma$$

$v \in H^1(\Omega)$

avec

$\Omega =]0,1[\times]0,1[$, $\Gamma = \delta\Omega$ = frontière de Ω ,

$f = 1$,

$\psi(v)$ primitive nulle pour $v = 0$ de $\Phi(v)$ avec

$$\Phi(v) = \begin{cases} g_o & \text{si } v|_\Gamma \geq kg_o \text{ , } g_o > 0 \text{ ,} \\ \frac{v}{k} & \text{si } -kg_1 < v|_\Gamma < kg_o \text{ , } k > 0 \text{ ,} \\ -g_1 & \text{si } v|_\Gamma \leq -kg_1 \text{ , } g_1 > 0 \text{ .} \end{cases}$$

On a donc

$$\psi(v) = \begin{cases} \frac{1}{2}\frac{v^2}{k} & \text{, si } -kg_1 < v|_\Gamma < kg_o \text{ ,} \\ g_o v - \frac{1}{2}kg_o^2 & \text{, si } v|_\Gamma \geq kg_o \text{ ,} \\ -g_1 v - \frac{1}{2}kg_1^2 & \text{, si } v|_\Gamma \leq -kg_1 \text{ .} \end{cases}$$

La solution de ce problème vérifie

$$\left|\begin{array}{ll} -\Delta u + u = f & \text{dans } \Omega \,, \\ \dfrac{\partial u}{\partial n} + \Phi(u) = 0 & \text{sur } \Gamma \,. \end{array}\right.$$

Le problème a été résolu pour

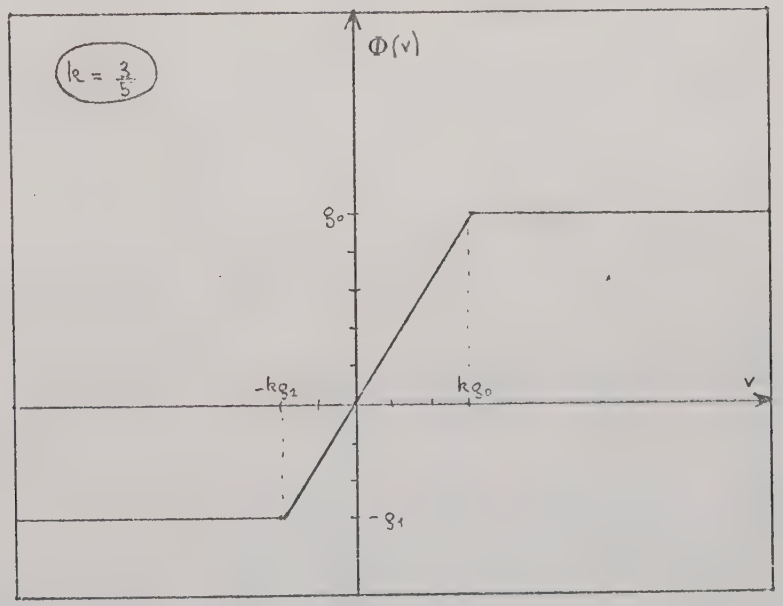
$$g = g_0 = g_1 = 0.3$$

et plusieurs valeurs de k :

$$k = 1,\ 0.1,\ 0.01,\ 0.001,\ \ldots$$

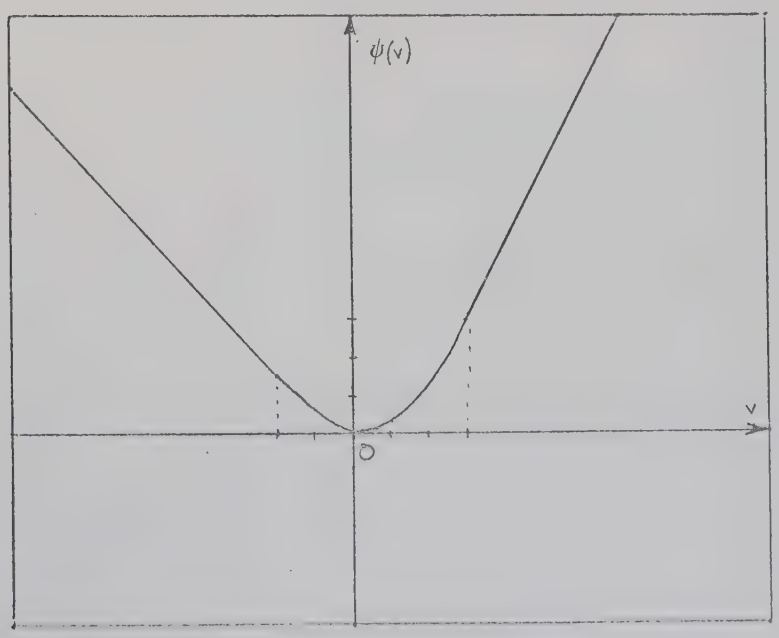
Graphe de $\Phi(v)$

Figure 15



Graphe de $\psi(v)$

Figure 16



a) Premier schéma

On considère la discrétisation suivante

$$\Omega =]0,1[\times]0,1[$$

$$h = \{h_1, h_2\} \ , \ h_1 = h_2 = h$$

$$R_h = \left\{ M \mid M = (m_1 h + \frac{h}{2} \ , \ m_2 h + \frac{h}{2}), \ m_1, m_2 \in \mathbb{Z} \right\}$$

$$\Omega_h = \left\{ M \mid M \in R_h, \quad \varpi_h^1(M) \cap \Omega \neq \phi \right\}$$

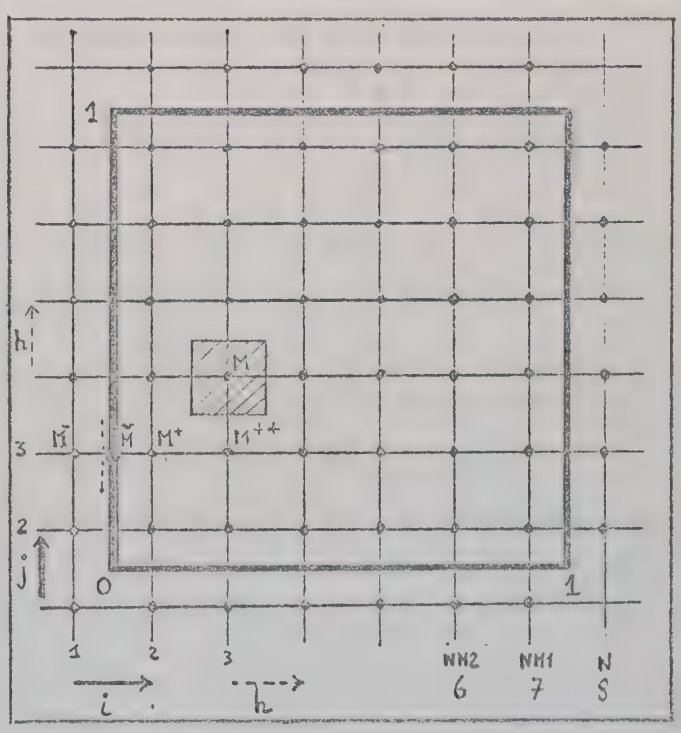
$$u_h = \sum_{M \in \Omega_h} u(M) . \theta_h^M \ , \ \theta_h^M \text{ fonction caractéristique de } \varpi_h^0(M)$$

$$\hat{\nabla}_i u_h = \frac{u_h(M + \frac{h}{2} \vec{e}_i) - u_h(M - \frac{h}{2} \vec{e}_i)}{h} \ , \ i = 1,2$$

$$J_h(v_h) = \frac{1}{2} \sum_{i=1}^{2} \int_\Omega |\hat{\nabla}_i v_h|^2 dx + \frac{1}{2} \int_\Omega v_h^2 dx - \int_\Omega f v_h dx + \int_\Gamma \psi_h(\tilde{v}_h) d\Gamma$$

$\tilde{v}_h$ étant une moyenne entre les valeurs voisines de v_h sur la frontière (cf. Figure 17) ; par exemple sur

Figure 17



$\{o\} \times [0,1]$

on a,

$$\tilde{v}_h(\tilde{M}) = \frac{v_h(M^+) + v_h(M^-)}{2} .$$

On définit, pour $x \in \Gamma$

$\tilde{v}_h(x) = v_h(\tilde{M})$

avec $\tilde{M}$ = multiple de $\frac{h}{2} \vec{e}_2$,

$|x - \tilde{M}| < \frac{h}{2}$.

On pose

$NM2 = \frac{1}{h}$,

$NM1 = NM2 + 1$,

$N = NM1 + 1$.

Le problème a été résolu pour

$h = \frac{1}{15}$, la fonctionnelle discrète comprend alors 289 variables ,

$h = \frac{1}{20}$, la fonctionnelle discrète ayant 484 variables.

Le test d'arrêt porte sur

$$\|J_h'(v_h)\|_1 = \sum_{M \in \Omega_h} \left| \frac{\partial J(v_h)}{\partial v_h(M)} \right| * \frac{1}{\text{nombre de variables}}$$

et

$$\|J_h'(v_h)\|_1 < 10^{-6} .$$

Dans les tableaux on trouvera en outre

$$\|J_h'(v_h)\|_2 = \sqrt{\sum_M \left| \frac{\partial J_h(v_h)}{\partial v_h(M)} \right|^2} .$$

Pour la minimisation on a utilisé la méthode du gradient conjugué adaptée au cas non quadratique.

b) Caractéristiques du problème en fonction de g et de k (premier schéma).

Pour $h = \frac{1}{15}$, avec pour solution de départ

$$v_h = 0.$$

on obtient les résultats consignés dans le tableau suivant (Tableau 8).
(L'ensemble des résultats de ce tableau a été obtenu en 1mn, 35 s 47 sur IBM-360-90).

Sur la Figure 18 on a représenté les valeurs de la fonction $u(x)$ sur $\frac{1}{4}$ du domaine $]0,1[x]0,1[$; les autres valeurs s'obtiennent par symétrie.

Sur la Figure 19 on a tracé quelques équipotentielles de $u(x)$ sur $\frac{1}{4}$ du domaine $]0,1[x]0,1[$, (le pas est $h = \frac{1}{20}$).

Sur la Figure 20 on donne la trace $u_h|_\Gamma$ de la solution u_h pour plusieurs valeurs de k, (la trace est prise = à $\tilde{u}_h$ comme défini précédemment), (le pas est $h = \frac{1}{20}$).

Tableau 8

g	k	Nombre d'itérations	Valeur de $J_h(v_h)$	$\|J'(v)\|_1$	$\|J'(v)\|_2$	plus grande val. u_h sur Ω
2	0.1	34	-0.02930	8.10^{-7}	18.10^{-6}	0.094
	0.01	43	-0.01834	8.10^{-7}	17.10^{-6}	0.073
	0.001	215	-0.01718	9.10^{-7}	21.10^{-6}	0.070
1	0.1	34	-0.02930	8.10^{-7}	18.10^{-6}	0.095
	0.01	43	-0.01834	8.10^{-7}	17.10^{-6}	0.073
	0.001	265	-0.17178	1.10^{-6}	32.10^{-6}	0.070
$\frac{1}{2}$	0.1	34	-0.02930	8.10^{-7}	18.10^{-6}	0.095
	0.01	43	-0.01834	8.10^{-7}	17.10^{-6}	0.073
	0.001	467	-0.17178	9.10^{-7}	19.10^{-6}	0.070
$\frac{1}{4}$	0.1	56	-0.02985	9.10^{-7}	20.10^{-6}	0.106
	0.01	107	-0.01975	8.10^{-7}	17.10^{-6}	0.094
	0.001	418	-0.01873	9.10^{-7}	22.10^{-6}	0.093
$\frac{1}{8}$	0.1	415	-0.13329	9.10^{-7}	17.10^{-6}	0.52
	0.01	312	-0.13048	9.10^{-7}	20.10^{-6}	0.52
	0.001	312	-0.13020	0.10^{-7}	20.10^{-6}	0.52
$\frac{1}{16}$	0.1	406	-0.28332	1.10^{-6}	21.10^{-6}	0.76
	0.01	403	-0.28262	9.10^{-7}	20.10^{-6}	0.76
	0.001	406	-0.28255	1.10^{-6}	21.10^{-6}	0.76
$\frac{1}{32}$	0.1	105	-0.38333	1.10^{-6}	23.10^{-6}	0.88
	0.01	105	-0.38316	1.10^{-6}	23.10^{-6}	0.88
	0.001	105	-0.38320	1.10^{-6}	23.10^{-6}	0.88
$\frac{1}{64}$	0.1	191	-0.43958	1.10^{-6}	20.10^{-6}	0.94
	0.01	191	-0.43954	1.10^{-6}	20.10^{-6}	0.94
	0.001	191	-0.43954	1.10^{-6}	20.10^{-6}	0.94

$u_h|_\Gamma = \tilde{u}_h$ est dans la "zone oblique" de $\Phi(v)$.

$u_h|_\Gamma = u_h$ est dans la "zone constante" de $\Phi(v)$.

Commentaires : $h = \frac{1}{\quad}$, $f = 1$, $\Omega =]0,1[\times]0,1[$

Figure 18

Valeurs de u(x) sur 1/4 du domaine
Les données sont f=1. , g=0.3 ,
k=1. , h=1/20 .
Les résultats ont été obtenus par
la méthode du gradient conjugué .

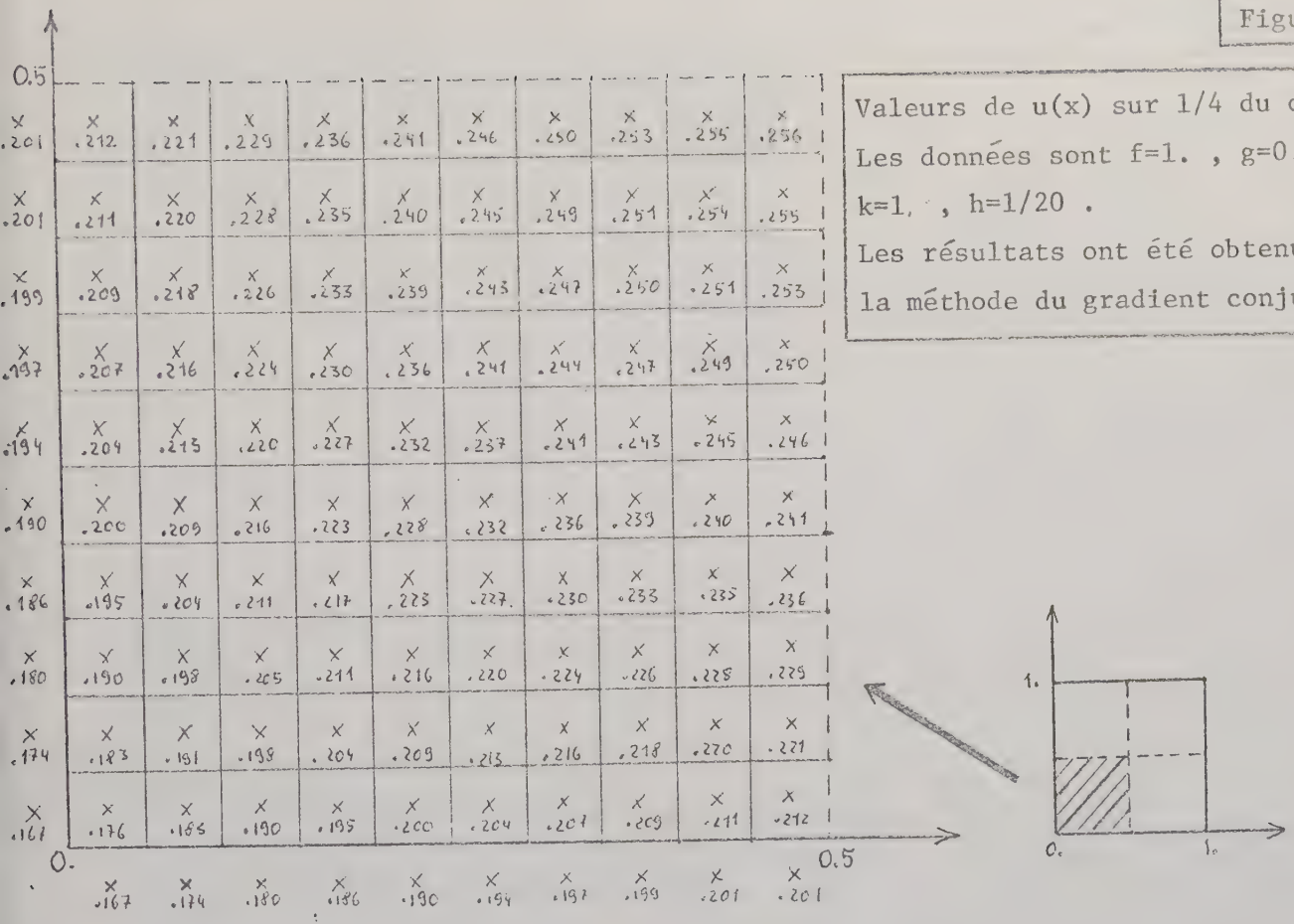


Figure 19

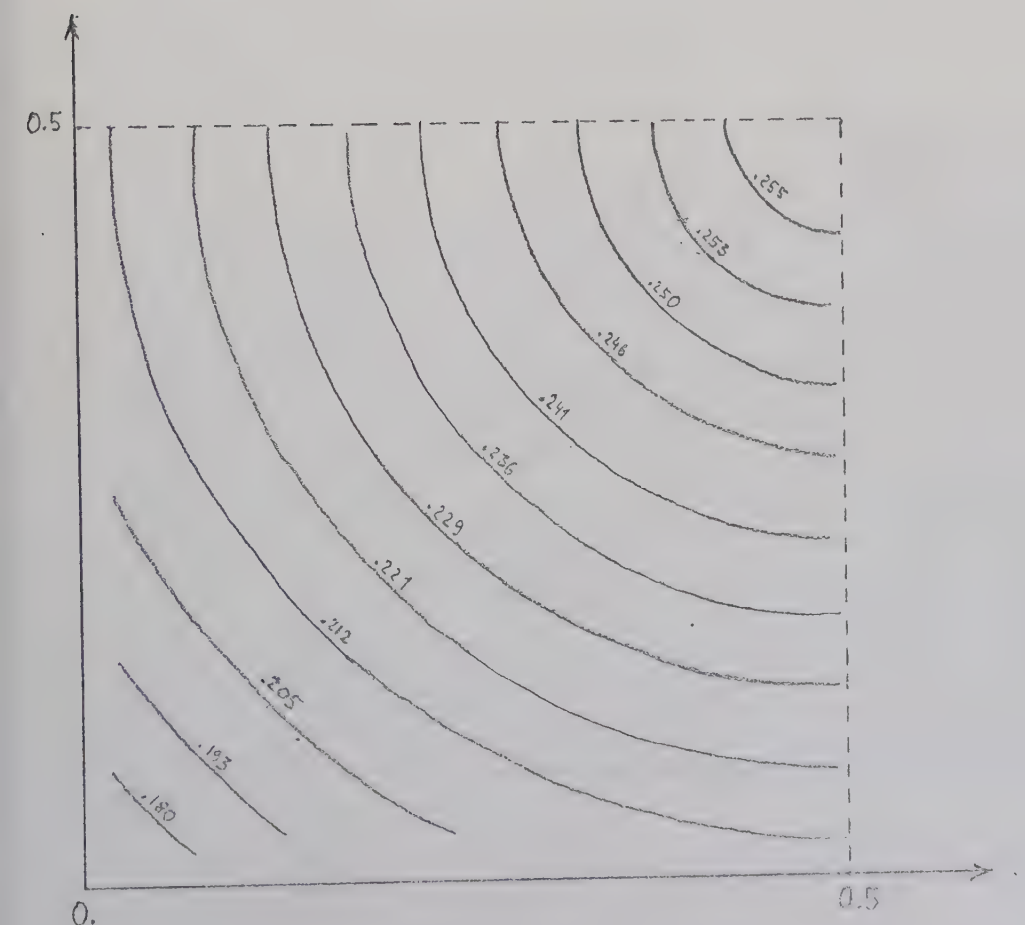


Figure 20

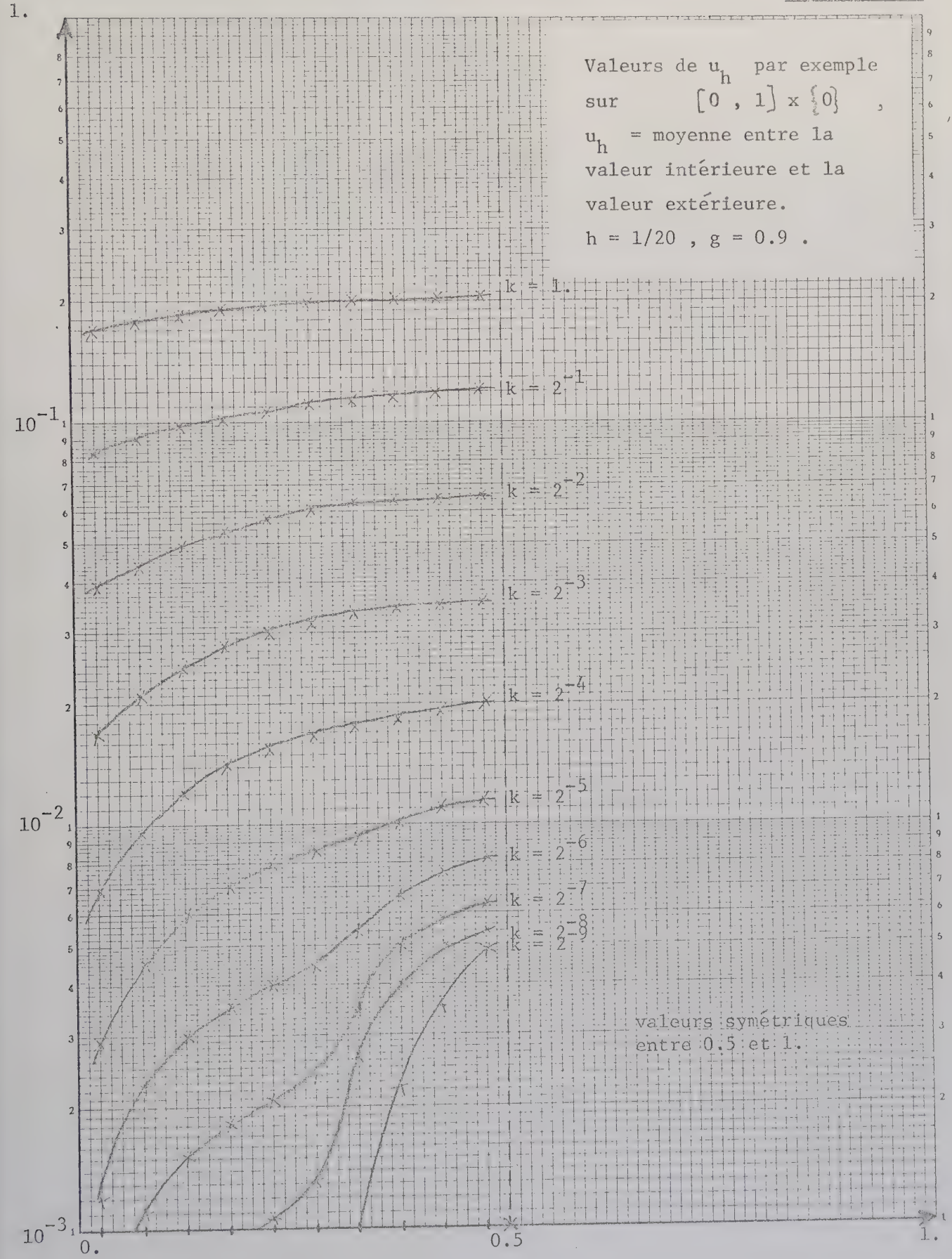
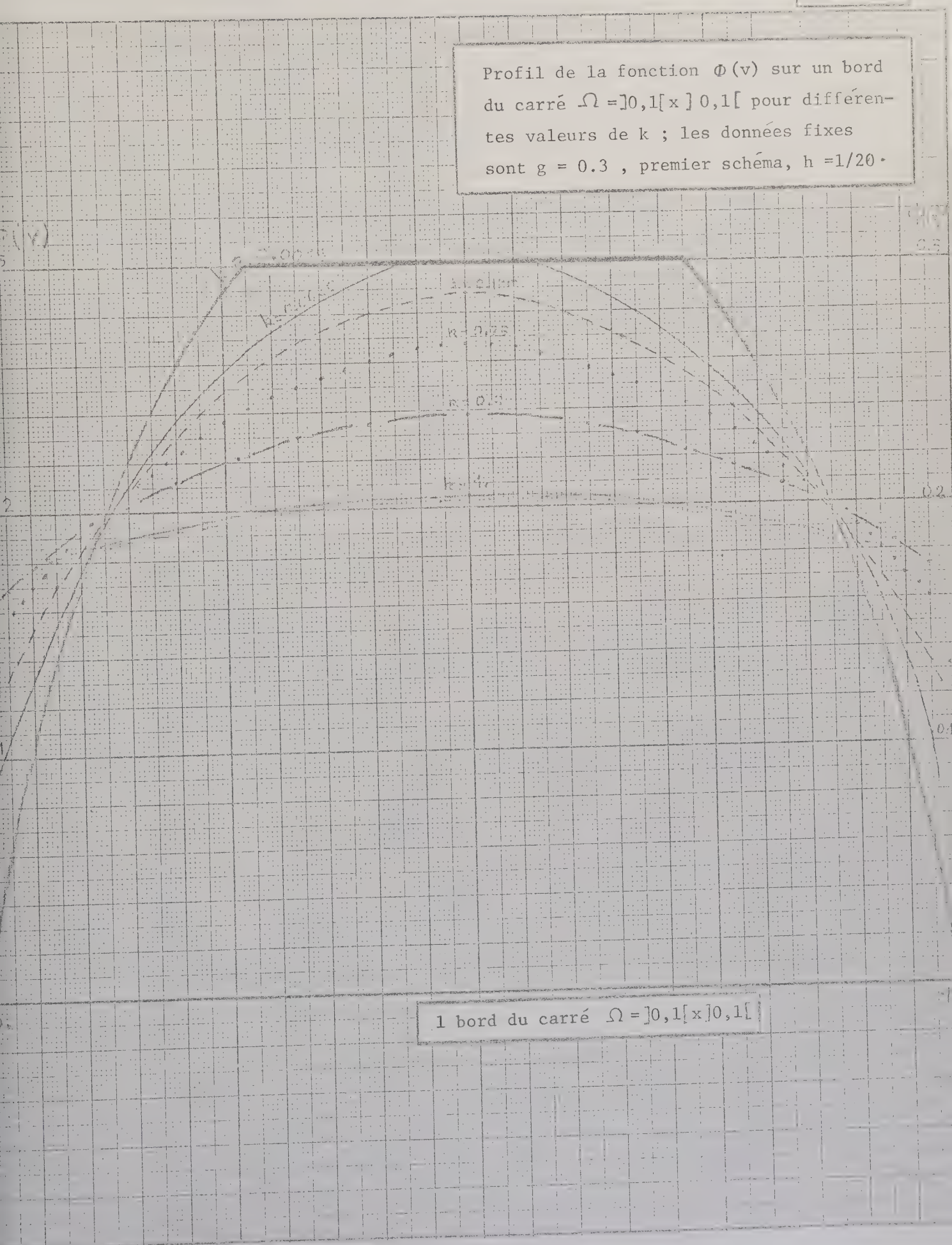

1.
10^{-1}
10^{-2}
10^{-3}
0.
0.5
1.
Valeurs de u_h par exemple sur $[0, 1] \times \{0\}$,
u_h = moyenne entre la valeur intérieure et la valeur extérieure.
h = 1/20 , g = 0.9 .
k = 1.
$k = 2^{-1}$
$k = 2^{-2}$
$k = 2^{-3}$
$k = 2^{-4}$
$k = 2^{-5}$
$k = 2^{-6}$
$k = 2^{-7}$
$k = 2^{-8}$
$k = 2^{-9}$
valeurs symétriques entre 0.5 et 1.

Figure 21



Profil de la fonction $\Phi(v)$ sur un bord du carré $\Omega =]0,1[\times]0,1[$ pour différentes valeurs de k ; les données fixes sont g = 0.3 , premier schéma, h =1/20.

1 bord du carré $\Omega =]0,1[\times]0,1[$

c) Deuxième schéma

On reprend les mêmes données que pour le schéma précédent, mais ici on résout directement le problème discrétisé par sur-relaxation. Les équations discrétisées s'écrivent

$$\begin{cases} -\Delta_h u_h + u_h = f_h & \text{sur } \Omega_h =]0,1[\times]0,1[, \\ \nu_h(u_h) + \Phi(u_h) = 0 & \text{sur } \Gamma , \end{cases}$$

où

Δ_h est une approximation de Δ , et
ν_h est une approximation de $\frac{\partial}{\partial n}$:

à savoir

$$\Delta_h u_h(x) = \sum_{i=1,2} \frac{u_h(x + h\vec{e}_i) - 2u_h(x) + u_h(x - h\vec{e}_i)}{h^2} ,$$

$\nu_h [u_h(x)]$ = valeur de la dérivée de la parabole passant par les trois points $u_h(M^-)$, $u_h(M^+)$, $u_h(M^{++})$; la dérivée étant prise en M^+ (cf. Figure 17, pour M^-, M^+, M^{++}).

On cherche u_h dans l'espace des fonctions

$\left\{ v_h \mid v_h = \sum_M v^M . \theta_h^M \ , \ M \in \widetilde{\Omega}_h \right\}$, θ_h^M fonction caractéristique de $\overline{\omega}_h^\circ(M)$

avec

$$\widetilde{\Omega}_h = \left\{ M \mid M \in \widetilde{R}_h \ , \ \overline{\omega}_h^1(M) \cap \widetilde{\Omega} \neq \phi \right\} ,$$

$$\widetilde{R}_h = \left\{ M \mid M = (m_1 h, m_2 h) \ , \ m_1 \ , \ m_2 \in \mathbb{Z} \right\} ,$$

(cf. Figure 22).

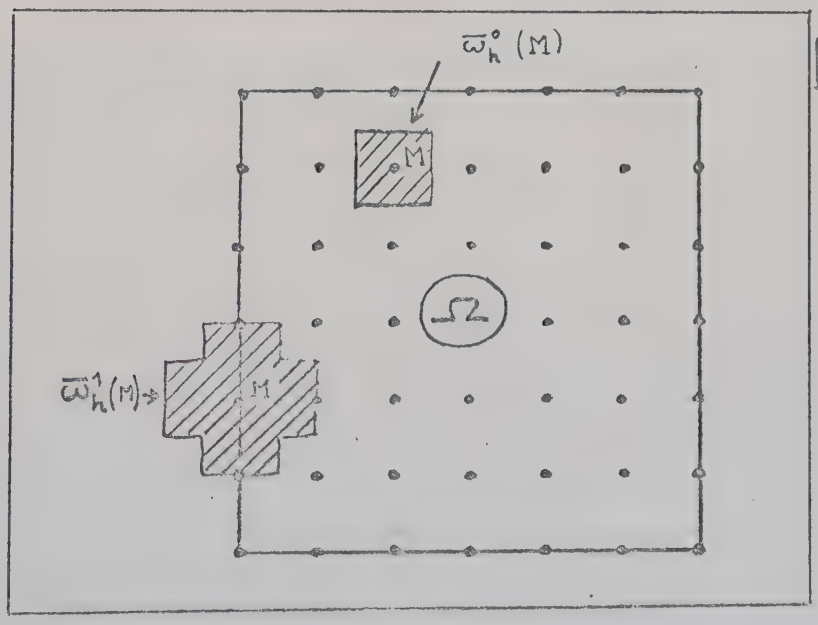


Figure 22

La méthode utilisée est la méthode de Gauss-Seidel avec une recherche de $\omega_{opt.expérimental}$ par tatonnements ; le point de départ est $v_h = 0$.

Avec un test d'arrêt portant sur

$$\sum_M \left| u^{n+1}(M) - u^n(M) \right| < 10^{-4}$$

on obtient le Tableau suivant (Tableau 9):

Tableau 9

g	k	Nombre d'itérations	$\omega_{opt.expérimental}$
1	0.1	46	1.77
	0.01	43	1.75
	0.001	42	1.75
$\frac{1}{4}$	0.1	76	1.85
	0.01	138	1.825
	0.001	336	1.625
$\frac{1}{16}$	0.1	199	1.9375
	0.01	200	1.9375
	0.001	200	1.9375

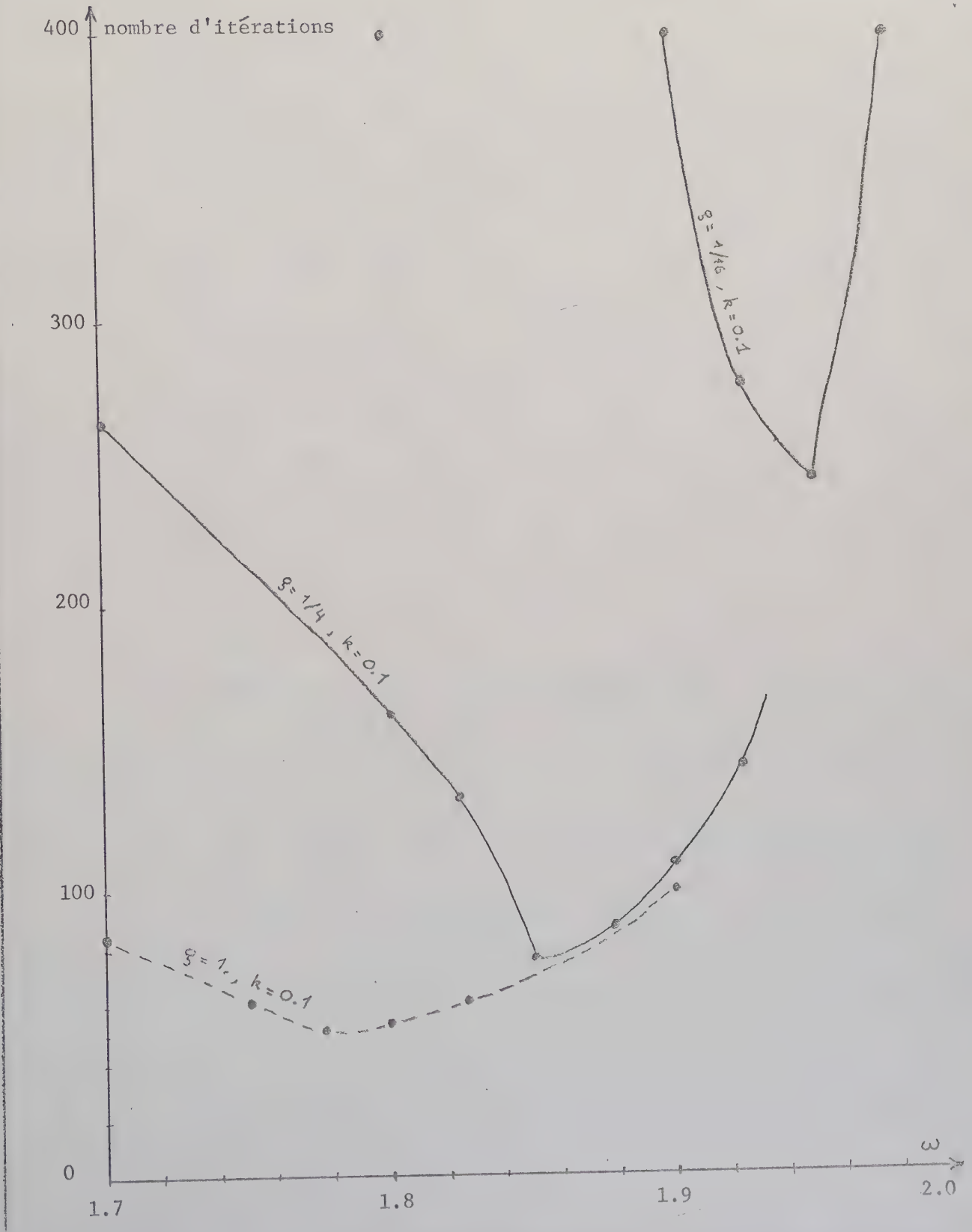
Nota 1 : Les résultats de ce tableau et du Tableau 8 ne doivent pas être comparés, les tests d'arrêt étant en effet tout à fait différents.

Nous donnons dans le graphique suivant la "sensibilité" de la méthode de Gauss-Seidel en fonction de ω (Figure 23); le test d'arrêt étant

$$\sum_{M} \left| u^{n+1}(M) - u^{n}(M) \right| < 10^{-4} :$$

Figure 23

Sensibilité de la méthode de Gauss-Seidel appliquée a la résolution d'un problème de climatisation en fonction de ω .
Test d'arrêt : $\sum_M |u^{n+1}(M) - u^n(M)| < 10^{-4}$.



2.4 Un problème non différentiable stationnaire

On considère le problème

$$\left|\begin{array}{l} \underline{\text{minimiser}}\ J(v) = \displaystyle\int_\Omega |\text{grad } v|^2\, dx - 2\int_\Omega fv\, dx + 2\int_\Omega g^+ v^+ dx \\ \qquad\qquad + 2\displaystyle\int_\Omega g^- v^- dx \\ v \in H_o^1(\Omega) \end{array}\right.$$

où

$$v^+ = \max\{0, v\}$$
$$v^- = \min\{0, v\}$$

et f, $g^+ \geqslant 0$, $g^- \leqslant 0$ sont donnés.

Ce problème peut encore s'écrire[1]

$$\left|\begin{array}{l} \underline{\text{minimiser}}\ J(v) = \displaystyle\int_\Omega |\text{grad } v|^2\, dx - 2\int_\Omega fv\, dx + 2\int_\Omega g_1 |v| dx \\ \qquad\qquad + 2\displaystyle\int_\Omega g_2\, v dx \\ v \in H_o^1(\Omega) \end{array}\right.$$

avec

$$g_1 = \frac{g^+ - g^-}{2}\ , \quad g_2 = \frac{g^+ + g^-}{2}$$

Nous prenons ici

$$\Omega =]0,1[$$

$$f = \begin{cases} 4\pi^2 \sin 2\pi x + g_1 & \text{si } x < \frac{1}{2} \\ 4\pi^2 \sin 2\pi x - g_1 & \text{si } x > \frac{1}{2} \end{cases}$$

$$g^+ = 12.5\ ,\ g^- = -12.5$$

$$g_1 = 12.5\ ,\ g_2 = 0.$$

Ce problème possède la solution unique

$$u = \sin 2\pi x\ . \quad \blacksquare$$

Le problème discrétisé avec les données ci-dessus et

$$h = \frac{1}{20}\ ,\ N = 19\ ,$$

$$v = \sum_{i=0}^{N+1} v_i \cdot \theta_h^i\ ,\ \theta_h^i \text{ fonction caractéristique de}$$

$$]\left(i - \tfrac{1}{2}\right)h,\ \left(i + \tfrac{1}{2}\right)h[$$

s'écrit

$$\left|\begin{array}{l} \underline{\text{minimiser}}\ J_h(v) = \displaystyle\sum_{i=0}^{N+1} \frac{(v_{i+1} - v_i)^2}{h^2} - 2f\sum_{i=0}^{N+1} v_i + 2g_1 \sum_{i=0}^{N+1} |v_i| \end{array}\right.$$

---

[1] Pour cela on note que $v^+ + v^- = v$, $v^+ - v^- = |v|$ et $2v^+ = v + |v|$, $2v^- = v - |v|$.

On pose

$$J_h(v) = J_{o,h}(v) + j_h(v)$$

avec

$$J_{o,h}(v) = \sum_{i=0}^{N+1} \frac{|v_{i+1} - v_i|^2}{h^2} - 2f \sum_{i=0}^{N+1} v_i \ ,$$

$$j_h(v) = \sum_{i=0}^{N+1} j_{i,h}(v_i) \quad , \quad j_i(v_i) = 2g_1 \cdot |v_i| \ . \ \blacksquare$$

a) Méthode de gradient

On se donne $\varepsilon > 0$, ε étant appelé à tendre vers 0, et l'on

$$\underline{\text{minimise}}\ \tilde{J}_h(v) = J_{o,h}(v) + \sum_{i=0}^{N+1} \max \left[j_{i,h}(v_i), \varepsilon \right] .$$

On utilise comme direction de descente

$$J'_{o,h}(v) + \sum_{i=0}^{N+1} d_i(v)$$

avec

$$d_i(v) = \begin{cases} j'_{i,h}(v) & \text{si } |v_i| \geqslant \varepsilon \ , \\ 0 & \text{sinon .} \end{cases}$$

On a utilisé la méthode de gradient à pas optimal. ■

b) Méthode d'approximation des termes non différentiables par d'autres fonctions

On remplace la fonction $x \longrightarrow |x|$ par une fonction $\omega(x)$ telle que

$$\begin{cases} \omega(x) = |x| & \text{si } |x| \geqslant \varepsilon > 0, \quad \text{fixe assez petit ,} \\ \omega(x) = F(x) & \text{si } |x| \leqslant \varepsilon \ , \end{cases}$$

où $F(x)$ est une fonction différentiable en 0, aussi "proche" que possible de $|x|$.

On étudie les fonctions :

- $F_1(x) = 1. - \cos mx$,
  le paramètre m étant choisi de façon que
  $1. - \cos mx = |x|$, pour $|x| = \varepsilon$.
- $F_2(x) = -\frac{1.}{2\varepsilon^3} * x^4 + \frac{3.}{2\varepsilon} * x^2$.
- $F_3(x) = x \ \text{th}(mx)$,
  le paramètre m étant choisi de façon que
  $x \ \text{th}(mx) = |x|$, pour $|x| = \varepsilon$.
- $F_4(x) = \sqrt{\eta + x^2}$,
  où η est un paramètre > 0, assez proche de 0 .
- $F_5(x) = |x|^{1+\eta}$,
  avec $\eta > 0$ assez petit.

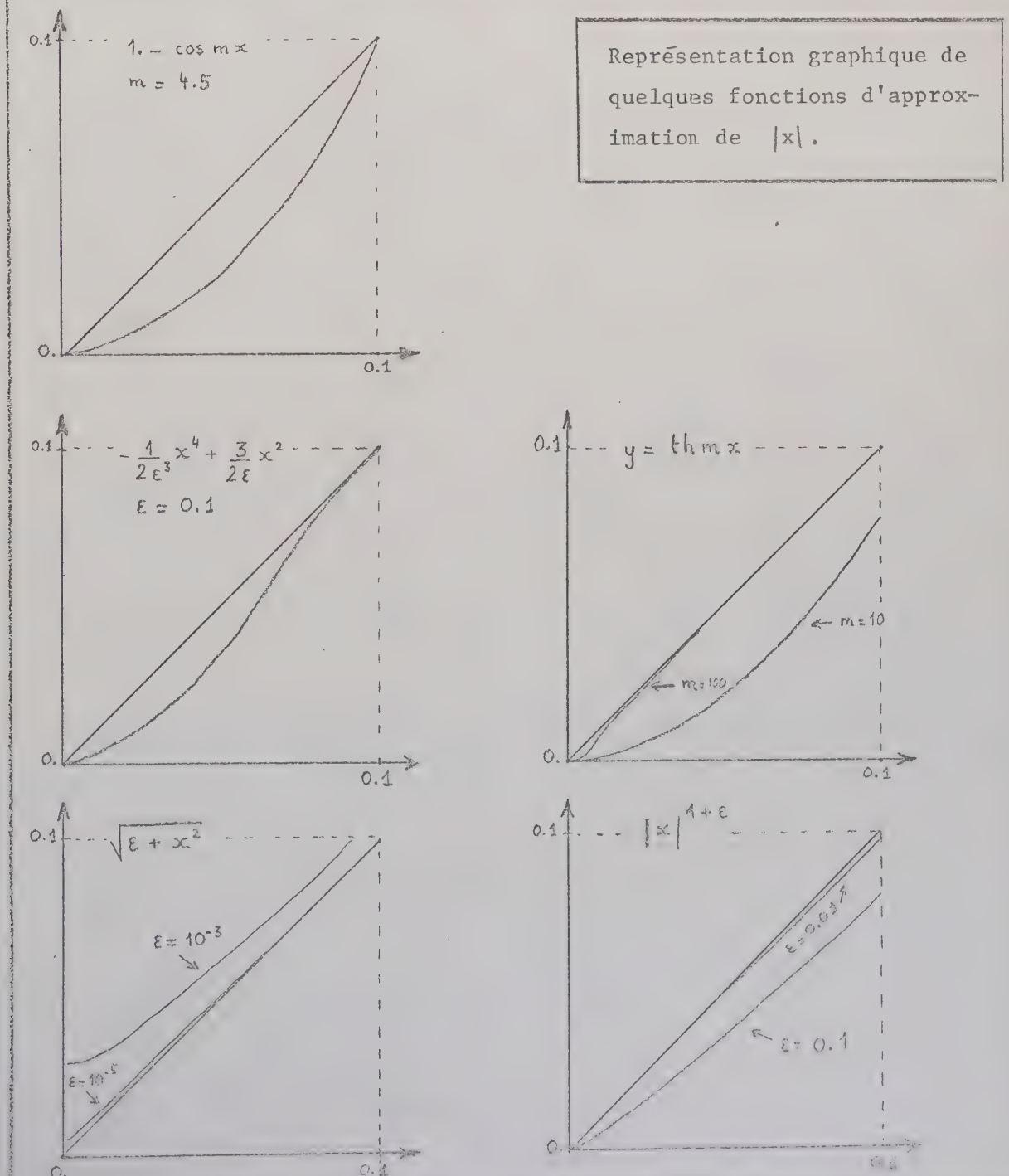
Avec les fonctions précédentes on a utilisé la méthode du gradient à pas optimal. ■

Une autre possibilité est de remplacer la fonction $x \longrightarrow |x|$ par une fonction $\omega(x)$ continuement différentiable et "proche" de $|x|$, par exemple

- $\omega_1(x) = \sqrt{\varepsilon + x^2}$,
- $\omega_2(x) = |x|^{1+\eta}$. ■

On a représenté sur le graphique suivant (Figure 24) les fonctions $F_1(x)$, ..., $F_5(x)$.

Figure 24



Représentation graphique de quelques fonctions d'approximation de $|x|$.

c) Méthodes de multiplicateurs

Ces méthodes reposent sur la remarque de CEA que

$$|x| = \max_{|q| \leq 1} qx$$

Posant $p = \{ p_o, \ldots, p_{N+1} \}$, et $K = \{ p \mid p \in \mathbb{R}^{N+2}, |p_i| \leq 1, i = 0, \ldots, N+1 \}$ le problème initial est alors équivalent à

$$\min_{v \in \mathbb{R}^{N+2}} \max_{p \in K} L(v,p)$$

avec

$$L(v,p) = J_o(v) + \sum_{i=0}^{N+1} g_1 \, p_i v_i \ .$$

On montre sans difficulté que le problème initial discrétisé est encore équivalent à

$$\max_{p \in K} \min_{v \in \mathbb{R}^{N+2}} L(v,p) \quad . \blacksquare$$

Ce problème est résolu par les deux méthodes suivantes :

1ère méthode (méthode de double gradient)

On se donne $v^o \in \mathbb{R}^{N+2}$, $p^o \in \mathbb{R}^2$ et l'on détermine, pour $n = 0,1,2,\ldots$

$$\left\| \begin{array}{l} v^{n+1} = v^n - \rho_v . L'_v(v^n, p^n) \ , \\ p^{n+1} = P_K \left[p^n + \rho_p \, L'_p(v^n, p^n) \right] \ , \end{array} \right.$$

où P_k désigne la projection sur K au sens suivant

$$P_k [p] = \left\{ \min \left[|p_o| , 1 \right] , \ldots, \min \left[|p_{N+1}| , 1 \right] \right\}.$$

Les paramètres ρ_v, ρ_p peuvent être pris fixes > 0; on dira alors qu'il s'agit d'une méthode de double gradient à pas constants.

Le paramètre ρ_p étant fixé > 0, on peut aussi choisir ρ_v de façon que

$$L(v^{n+1}, p^n) \leq L(v^n - L'_v(v^n, p^n), p^n), \quad \forall \rho \in \mathbb{R} .$$

On dira alors qu'il s'agit d'une méthode de double gradient à pas optimal en v. La valeur de ρ_v se calcule alors sans difficulté. ■

2e méthode (méthode de minimax semi-implicite)

On se donne p^o, et l'on détermine, pour $n = 0,1, \ldots$

$$\left\| \begin{array}{l} v^{n+1} \text{ tel que } L(v^{n+1}, p^n) \leq L(v, p^n), \ \forall v \in \mathbb{R}^{N+2} \ , \\ p^{n+1} = P_K \left[p^n + \rho_p \, L'_p(v^n, p^n) \right] \ , \end{array} \right.$$

où ρ_p est une constante > 0.

Pour un choix convenable de ρ_p la méthode est convergente.

Le calcul de v^{n+1} sera fait par les 3 méthodes :

- méthode du gradient à pas ajusté ,
- méthode du gradient à pas optimal ,
- méthode du gradient conjugué.

d) Comparaison des méthodes

Le point de départ commun à toutes les méthodes est

$$v_i = 0 \ , \ i = 0, \ \ldots \ N+1 \ ; \ (N = 19, \ h = \frac{1}{20}) .$$

On a retenu pour critère d'arrêt

$$\| "J_h'(v)" \|_2 \leqslant 0.01 \qquad (\text{si } v_i = 0, \text{ on pose } \frac{d|v_i|}{dv_i} = 0).$$

Avec ce critère on trouve

$$J_h(v) = 19.9017 \ .$$

Les principaux résultats sont consignés dans les 3 tableaux suivants (Tableaux 10, 11, 12).

Les calculs ont été effectués sur CII-10070 et les temps sont donnés en secondes.

Tableau 10

Méthodes	Temps	Nbre d'itérations	Caractéristiques	Commentaires
gradient: pas divisé ($\varepsilon = 0$)	0.052	150	--	--
gradient: pas optimal ($\varepsilon = 0$)	0.063	125	--	--
$F_1(x) = 1. - \cos(mx)$	0.042	130	$\varepsilon = 10^{-3}$, $m = 4$	Pour $\varepsilon = 0.1$, m les résultats sont pas satisfaisants
$F_2(x) = -\frac{1}{2\varepsilon^3} * x^4 + \frac{3.}{2\varepsilon} * x^2$	0.042	130	$\varepsilon = 10^{-3}$	Pour $\varepsilon = 0.1$ le résultats ne sont pas satisfaisants
$F_3(x) = x\,\mathrm{th}(mx)$	0.044	130	$\varepsilon = 10^{-1}$, $m = 100$	Pour $\varepsilon = 0.1$, m= les résultats sont pas satisfaisants
$F_4(x) = \sqrt{\eta + x^2}$	0.042	130	$\varepsilon = 10^{-1}$, $\eta = 10^{-7}$	Pour $\eta \geq 10^{-5}$ les résultats sont pas satisfaisants
$F_5(x) = \lvert x \rvert^{1+\eta}$	0.046	140	$\varepsilon = 10^{-3}$, $\eta = 0.1$	Résultats non satisfaisants pour $\varepsilon = 10^{-1}$, $\eta = 0.1$; $\varepsilon = 10$ $\eta = 10^{-5}$.

Tableau 11

METHODE DE DOUBLE GRADIENT			
Caractéristiques		Temps	Nombre d'itérations
ρ_v =0.005	ρ_p = 100	0.06	300
	ρ_p = 10	0.064	320
	ρ_p = 1	0.07	350
	ρ_p = 0.1	0.12	600
ρ_v =0.009	ρ_p = 100	0.028	140
	ρ_p = 10	0.034	170
	ρ_p = 1	0.05	225
	ρ_p = 0.1	0.07	350
ρ_v optimal	ρ_p = 100	0.082	275
	ρ_p = 10	0.102	340
	ρ_p = 1	0.103	345
	ρ_p = 0.1	0.105	350

Tableau 12

METHODE DE MINIMAX SEMI-IMPLICITE		
Caractéristiques	Temps	Nombre d'itérations
ρ_p = 100.	0.027	9
ρ_p = 10.	0.027	9
ρ_p = 1.	0.042	14
ρ_p = 0.1	0.042	14

Commentaires : Les 3 méthodes de gradient conjugué, de Gradient à pas divisé, et de gradient à pas optimal donnent sensiblement les mêmes résultats en nombre d'itérations et en temps. Les résultats donnés correspondent à la méthode du gradient conjugué utilisée pour la minimisation à chaque itération.

Par rapport aux autres méthodes, la méthode de minimax semi-implicite est, pour un bon choix de ρ_p, la plus rapide [1].

2.5 Le problème de la torsion élasto-plastique

On considère le problème

$$(I)\quad \left|\begin{array}{l} \underline{\text{minimiser}}\ J(v) = \frac{1}{2}\int_\Omega |\text{grad } v|^2\, dx - \int_\Omega f.v\, dx \\ \text{sous les contraintes} \\ v \in K = \{\, w \mid w \in H_o^1(\Omega),\ |\text{grad } w|^2 \leqslant 1.,\ \text{p.p.sur } \Omega \,\}, \\ \Omega \text{ étant un ouvert borné de } \mathbb{R}^2 \text{ et } f \text{ donné dans } L^2(\Omega). \end{array}\right.$$

On montre [2], si Ω est simplement convexe et si f = constante sur Ω, que ce problème est équivalent au suivant :

$$(II)\quad \left|\begin{array}{l} \underline{\text{minimiser}}\ J(v) = \frac{1}{2}\int_\Omega |\text{grad } v|^2\, dx - \int_\Omega f.v\, dx \\ \text{sous les contraintes} \\ v \in K = \{\, w \mid w \in H_o^1(\Omega),\ |v(x)| \leqslant d(x,\Gamma),\ \text{p.p. sur } \Omega \,\}, \\ d(x,\Gamma) \text{ étant la distance du point } x \in \Omega \text{ à la frontière } \Gamma \text{ de} \end{array}\right.$$

Nous nous plaçons ici dans le cas où

$$\Omega =]0,1[\times]0,1[,$$

$$h = h_1 = h_2 = \frac{1}{N+1} .$$

Nous allons donner des résultats concernant

- la résolution de (I) par la méthode des centres à troncature variable,
- La résolution de (II) par la méthode de Gauss-Seidel en projection et la méthode S.N.

a) Résolution de (I) par la méthode des centres à troncature variable

On considère la discrétisation classique de Ω dans le cas de $H_o^1(\Omega)$: R_h, Ω_h, V_h, v_h, $\nabla_h v_h$ comme au chapitre II, § 2.1 et 2.2. Le problème (I) discrétisé s'écrit :

$$\left|\begin{array}{l} \underline{\text{minimiser}}\ J_h(v_h) = \frac{1}{2}\int_\Omega |\nabla_h v_h|^2\, dx - \int_\Omega f.v_h\, dx \\ \text{sous les contraintes} \\ v_h \in K_h = \{\, w_h \mid w_h \in V_h,\ |\nabla_h w_h|^2 \leqslant 1,\ \text{p.p. sur } \Omega \,\}. \end{array}\right.$$

(1) Mais nous n'avons pas programmé la méthode du gradient conjugué adaptée au problème initial.

(2) cf. BREZIS-SIBONY [1].

En notant

$$v_{i,j} = v_h(M) \qquad \text{avec } M = \{ih,\ jh\}$$

le problème peut s'écrire

$$\underline{\text{minimiser}}\ J_h(u_h) =$$

$$= \sum_{i,j=1}^{N} \left\{ \frac{1}{4} \left[|u_{i+1,j} - u_{i,j}|^2 + |u_{i,j} - u_{i-1,j}|^2 + \right.\right.$$

$$+ |u_{i,j+1} - u_{i,j}|^2 + |u_{i,j} - u_{i,j-1}|^2 -$$

$$\left.\left. - f.u_{i,j}\ h^2 \right\}\right.$$

sous les contraintes, pour $i,j = 0, \ldots N+1$,

$$|u_{i+1,j} - u_{i,j}|^2 + |u_{i,j} - u_{i,j-1}|^2 \leqslant h^2 ,$$

$$|u_{i+1,j} - u_{i,j}|^2 + |u_{i,j+1} - u_{i,j}|^2 \leqslant h^2 ,$$

$$|u_{i,j} - u_{i-1,j}|^2 + |u_{i,j+1} - u_{i,j}|^2 \leqslant h^2 ,$$

$$|u_{i,j} - u_{i-1,j}|^2 + |u_{i,j} - u_{i,j-1}|^2 \leqslant h^2 ,$$

et $u_{i,j} = 0$ si $i = 0$ ou $N+1$, ou si $j = 0$ ou $N+1$.

On vérifiera sans peine que le problème ainsi écrit comprend

N^2 variables (noter que $u_{o,j}$, $u_{N+1,j}$, $u_{i,o}$, $u_{i,N+1}$ sont imposés égaux à 0),

$4(N^2 + N)$ contraintes différentes et non trivialement vérifiées.

On a pris $f = 10$.

Nous avons résolu ce problème par la méthode des centres à troncature variable avec pour fonction distance

$$D(v,\lambda) = \left[\lambda - J_h(v) \right]^p . \prod_{i \in I} G_i(v)$$

où I est l'ensemble $\{1,2, \ldots M\}$, M étant le nombre de contraintes, ces contraintes étant écrites sous la forme

$$G_j(v) \geqslant 0 .$$

L'extrapolation, i.e. la détermination de w^n (cf. ce même chapitre, § 1.4) est faite pour les deux procédés signalés à la Remarque 1.5 ; on prend pour w^n la meilleure valeur trouvée.

La maximisation requise pour déterminer chaque centre est faite par la méthode du gradient conjugué adaptée au cas non linéaire avec recherche du pas optimal par dichotomie.

L'exposant p intervenant dans $D(v,\lambda)$ est pris variable et sera précisé ultérieurement.

La troncature ρ est prise égale à 0.95.

Le point de départ est $v_h = 0$.

On donne les principaux résultats dans le Tableau 13 ci-après. ■

Tableau 13

h	$J_h(v)$ à l'optimum obtenu	Nombre de variables	Nombre de contraintes	Temps[1]
$\frac{1}{10}$	-2.360095	81	360	24 sec.
$\frac{1}{20}$	-2.396686	361	1520	2mn.19
$\frac{1}{30}$	-2.402672	841	3480	6mn.25
$\frac{1}{50}$	-2.404895	2401	9800	14mn.41[2]

Commentaires : [1] Les programmes ont été résolus sur IBM-360-91.

[2] Arrêt sur temps limite.

Dans le Tableau suivant (Tableau 14) nous donnons le détail de la convergence pour $h = \frac{1}{10}$. ■

tableau 14

Numéro du centre $v(\lambda_h)$	Nombre d'itérations retenues pour le calcul de $v(\lambda_n)$	Valeur[0] de $J_h(v(\lambda_n))$	Val.[1] de l'extrapolation $J_h(w(\lambda_n)$	Val.[2] de l'extrapolation $J_h(w(\lambda_n)$	Valeur de p[3] retenue
1	5	1.0226	1.6024	1.6868	27
2	5	1.7872	1.9466	2.2129	13.5
3	9	2.1981	2.2417	2.3045	6.75
4	5	2.3013	2.3171	2.3309	3.37
5	7	2.3296	2.3354	2.3434	1.68
6	6	2.3428	2.3449	2.3463	1.
7	11	2.3462	2.3479	2.3473	1
8	11	2.3478	2.3491	2.3497	1
9	13	2.3497	2.3503	2.3518	1.
10	11	2.3518	2.3527	2.3529	1.
11	11	2.3529	2.3538	2.3549	1.
12	11	2.3549	2.3556	2.3568	1.
13	11	2.3567	2.3573	2.3581	1.
14	12	2.3581	2.3584	2.3586	1.
15	12	2.3586	2.3588	2.3589	1.
...	...	...	...	...	...

Commentaires tableau 14

(o) Les chiffres donnés correspondent à - J(v).

(1) Extrapolation sur - $J'(v(\lambda_n))$, i.e. le premier procédé.

(2) Extrapolation sur la ligne des deux derniers centres.

(3) La formule est $p^n = \max\left[\frac{p^{n-1}}{2}, 1.\right]$ $p^o = \frac{N}{3}$.

Sur la Figure 25 on a représenté les zones de saturation du convexe.

La discrétisation précédente conduit à tenir compte d'un très grand nombre de contraintes, puisqu'environ 4 fois le nombre de variables. En adoptant une autre discrétisation de K on peut réduire ce nombre. La discrétisation de

$$K = \{ w \mid w \in H_o^1(\Omega), \quad |\text{grad } w|^2 \leq a, \text{ p.p. sur } \Omega \}$$

que l'on peut retenir est, pour i,j = 0,..., N+1,

$$|u_{i+1,j} - u_{i-1,j}|^2 + |u_{i,j+1} - u_{i,j-1}|^2 \leq a.(2h)^2.$$

Il y a alors $(N+2)^2$ contraintes pour N^2 variables.
Nous avons résolu le problème correspondant avec a = 2.
(dans ce qui précède nous avions pris a = 1).

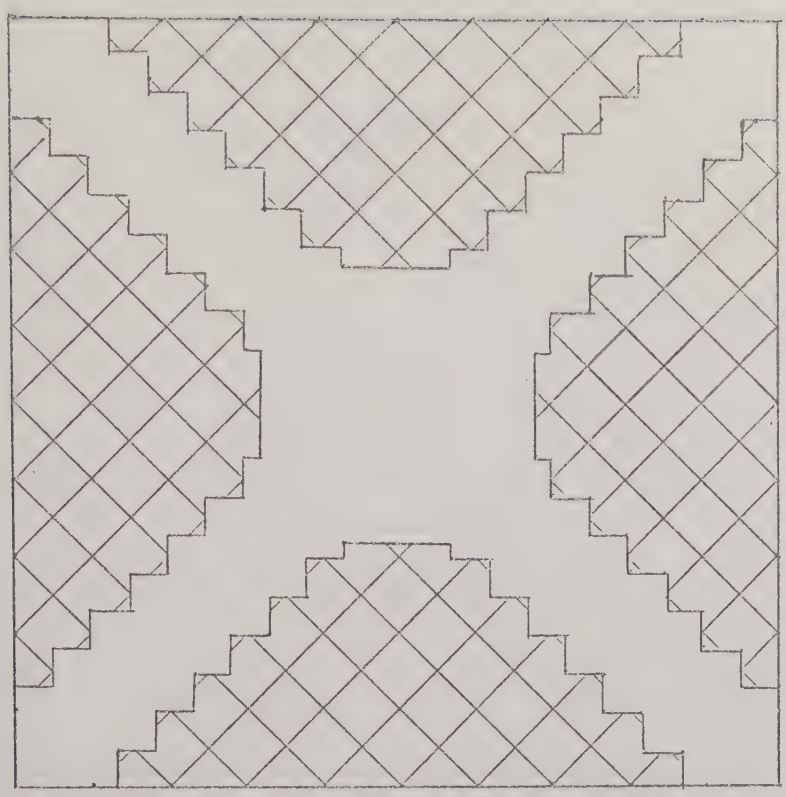
Pour différentes valeurs de h on obtient les principaux résultats figurés dans le Tableau 15:

Tableau 15

h	J(v) à l'optimum obtenu	Nombre de variables	Nombre de contraintes	Temps IBM-360-91
$\frac{1}{10}$	-2.87653	81	121	13 sec
$\frac{1}{20}$	-2.92321	361	441	1mn.46
$\frac{1}{30}$	-2.92913	841	961	5mn.08

Les zones de "plasticité" ($|\text{grad } u|^2 = 2.$) sont représentées sur la Figure 26.

Figure 25



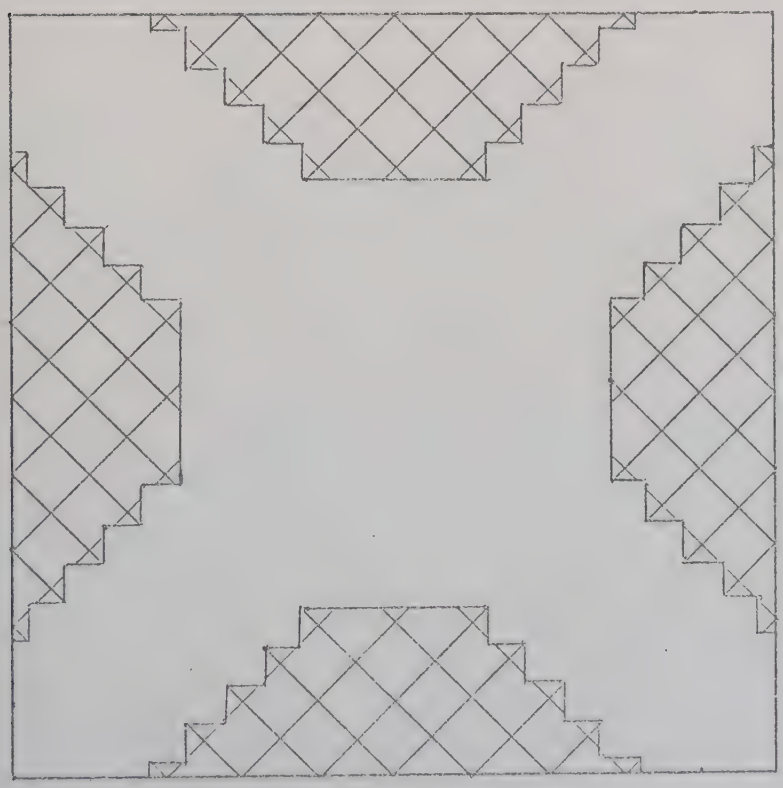
Résolution du problème

$$\min \frac{1}{2}\int_\Omega |\text{grad}\, v|^2 dx - 10.\int_\Omega v\, dx$$

$$v \in K = \{w \mid w \in H_0^1(\Omega),\ |\text{grad}\, w|^2 \leq 1,\ \text{p.p.}\}$$

Les résultats sont donnés pour $h = \frac{1}{20}$, d'où 361 variables et 1520 contraintes. Le problème est résolu par la méthode des centres à troncature variable. La zone hachurée correspond à $|\nabla_h v_h|^2 = 1.$ à 10^{-2} près.

Figure 26



Résolution du problème

$$\min \frac{1}{2}\int_\Omega |\text{grad}\, v|^2 dx - 10.\int_\Omega v\, dx$$

$$v \in K = \{w \mid w \in H_0^1(\Omega),\ |\text{grad}\, v|^2 \leq 2,\ \text{p.p.}\}$$

Les résultats sont donnés pour $h = \frac{1}{20}$, d'où 361 variables et 441 contraintes. Problème résolu par la méthode des centres à troncature variable. La zone hachurée correspond à $|\nabla_h v_h|^2 = 2.$ à 10^{-2} près.

b) Résolution de (II) par les méthodes de Gauss-Seidel en projection et la méthode S.N.

On cherche à résoudre le problème II sous la forme discrète

$$\underline{\text{minimiser}}\ J_h(v_h) =$$

$$= \sum_{i,j=1}^{N} \left\{ \frac{1}{4} \left[|v_{i+1,j} - v_{i,j}|^2 + |v_{i,j} - v_{i-1,j}|^2 + \right.\right.$$

$$\left.\left. + |v_{i,j+1} - v_{i,j}|^2 + |v_{i,j} - v_{i,j-1}|^2 \right] - f.v_{i,j}h^2 \right\}$$

sous les contraintes, pour $i,j = 1,\ldots, N$

$$|v_{i,j}| \leq \min\left[ih, jh, 1 - ih, 1 - jh\right].$$

En fait, la solution u_h étant positive on se restreint aux contraintes

$$v_{i,j} \leq \min\left[ih, jh, 1 - ih, 1 - jh\right].$$

On a pris successivement f = 1, 5, 10 (pour 1 le problème est en fait "sans contrainte", aucune contrainte n'étant saturée). Ce problème a été résolu par

- la méthode de sur-relaxation en projection (G.S.P.),
- la méthode S.N. adaptée (S.N.).

La modification de la méthode S.N. considérée est la suivante :

Etant donné d^n comme au Ch. VII, 1., Remarque 1.6, on définit

$$v^{n+1} = P_K\left[v^n + \rho^n d^n\right]$$

où ρ^n vérifie $J(v^n + \rho^n d^n) \leq J(v^n + \rho d^n)$, $\forall \rho$, ($v^n + \rho^n d^n$ peut donc être à l'extérieur du domaine K des contraintes).

On a pris $h = \frac{1}{40}$.

Les principaux résultats comparatifs entre les deux méthodes précitées sont portés dans le Tableau 16 ci-après.

Tableau 16

f	Méthode G.S.P.			Méthode S.N.		Test d'arrêt commun portant sur $\|\mathcal{E}^N J'_h(v_h)\| \leq$ valeurs ci-dessous, et valeur de u au centre de Ω.
	T	N	ω	T	N	
1.	1.117	201	1.8	0.567	59	10^{-7} ; .073323
	0.714	128	1.8	0.467	48	10^{-5} ; ~ 0.73295
	1.146	213	1.7	--	--	
5.	0.584	102	1.8	0.782	83	10^{-8} ; .324489
10.	0.347	86	1.8	0.821	88	10^{-8} ; .409879 (1)

Commentaires :

T = temps en minutes sur CII-10070.

N = nombre d'itérations pour satisfaire le test d'arrêt; on entend par nombre d'itérations le nombre de fois qu'on a déterminé de nouvelles directions de descente.

ω = paramètre de sur-relaxation utilisé dans la méthode G.S.P.; la valeur $\omega = 1.8$ est la meilleure valeur expérimentale pour f = 5. et f = 10. obtenue en faisant les essais 1.1, 1.2, ..., 1.9.

(1) avec les mêmes données un passage sur IBM-360-91 a donné T = 1.14 s avec pour test d'arrêt $0.8.10^{-3}$.

c) Conclusions

On notera l'intérêt des méthodes G.S.P. et S.N. par rapport à une méthode de pénalisation telle que la méthode des centres sur-relaxée, puisqu'en effet les temps de calcul sont de l'ordre de la seconde pour les premières et d'environ 10 minutes pour cette dernière.

Quant aux méthodes G.S.P. et S.N. elles sont sensiblement équivalentes en temps, la première étant d'autant plus rapide et la seconde d'autant moins rapide qu'il y a plus de contraintes.

2.6 Le problème de la torsion élasto-plastique en évolution

On considère l'inéquation parabolique, pour $t \in [0,T]$,

$$(u' - \Delta u - f,\ v - u) \geq 0, \quad \forall v \in K, \quad u \in \mathcal{K},$$

avec

$$K = \left\{ v \mid v \in H_o^1(\Omega),\ |v(x)| \leq d(x,\Gamma),\ \text{p.p. sur } \Omega \right\},$$

$$\mathcal{K} = \left\{ v \mid v \in W_c(0,T),\quad v \in K,\ \text{p.p. en } t \in [0,T] \right\}.$$

On prend $\Omega =]0,1[$ et l'intervalle de temps $[0,1]$.

On pourra vérifier que pour

$$f(x,t) = \begin{cases} 100 \quad \text{si} \quad x < \frac{t}{2}, \text{ ou si } x > \frac{t}{2}, \\ -\frac{1}{(t-1)^2}\left(x^2 - x - \frac{t(t-2)}{4}\right) - \frac{2}{t-1} \quad \text{si} \quad \frac{t}{2} \leq x \leq 1 - \frac{t}{2} \end{cases}$$

et

$$u_o(x) = x(1-x)$$

le problème a la solution exacte

$$u(x,t) = \begin{cases} x \quad \text{si} \quad 0 \leq x \leq \frac{t}{2} \\ \frac{1}{t-1}\left[x^2 - x + \frac{t^2}{4}\right] \quad \text{si} \quad \frac{t}{2} < x < 1 - \frac{t}{2} \\ 1 - x \quad \text{si} \quad 1 - \frac{t}{2} \leq x \leq 1 \end{cases} .$$

La discrétisation en espace est la discrétisation habituelle de $H_o^1(\Omega)$.

Des essais ont été conduits avec $h = \frac{1}{5}, \frac{1}{10}, \frac{1}{20}$ (soit un nombre de variables $N = \frac{1}{h} - 1$).

Le pas de discrétisation en temps est $k = \frac{1}{NT}$, où NT est le nombre d'intervalles de temps considérés.

La valeur des solutions approchée est mesurée à l'instant n.k par

$$ECART_2 = \left[\sum_{i=1}^{N} h. \left| u(ih,nk) - u_h(ih,nk) \right|^2 \right]^{1/2}$$

Dans tout ce paragraphe les résultats ont été obtenus sur BULL-GE-265 et les temps donnés en 1/6e de seconde.

a) Evolution de la solution

L'évolution de u(x,t) en fonction de t = 0 à 1 est représentée sur le graphique ci-après (Figure 27).

Dans ce qui suit nous allons étudier

- le schéma explicite,
- le schéma implicite en équation,
- le schéma implicite en projection.

Le schéma implicite en équation a été résolu par

- la méthode de Gauss-Seidel avec paramètre optimal,
- la méthode du gradient conjugué.

Le schéma implicite en projection a été résolu par

- la méthode de Gauss-Seidel "projeté" avec paramètre optimal,
- la méthode S.N.

Par "paramètre optimal" nous entendons la valeur théorique optimale du paramètre ω pour le cas sans contraintes.

b) Etude du schéma explicite

Avec

$$\left|\begin{array}{l} u^n = \sum_{i=1}^{N} u_i^n \cdot \bar{\omega}_h^o(ih) \quad , \quad N = \frac{1}{h} - 1 \ , \ n = 1, \ldots, NT, \\ u_i^o = ih(1 - ih) \ , \\ f_i^n = f(ih, nk) \ , \\ d_i = d(ih, \Gamma) \ . \end{array}\right.$$

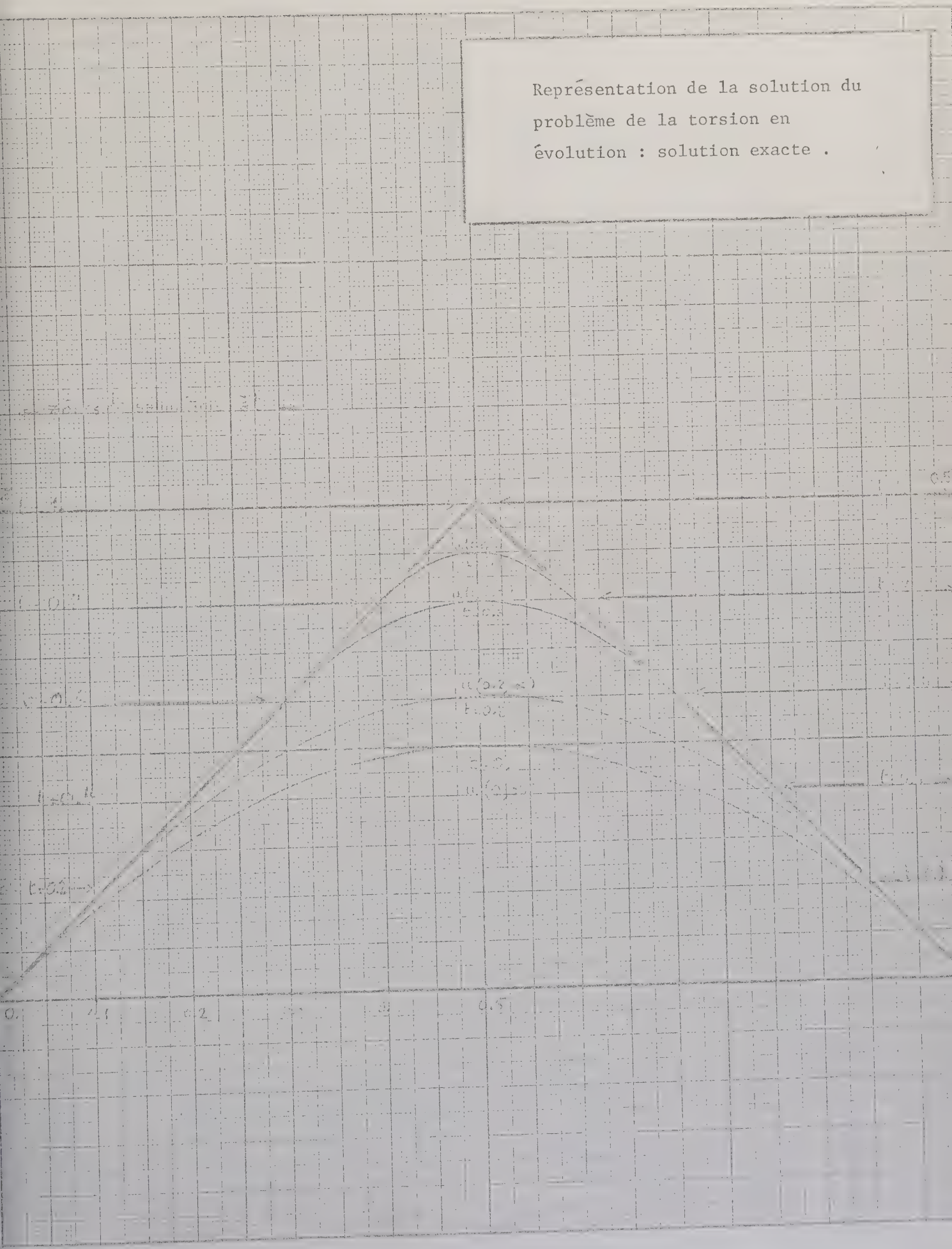
Le schéma s'écrit

$$\left\|\begin{array}{l} u_1^{n+1} = \underline{\min} \left[\left| u_1^n + \frac{k}{h^2} (-2u_1^n + u_2^n) + k f_1^{n+1} \right| , d_1 \right] , \\ u_i^{n+1} = \underline{\min} \left[\left| u_i^n + \frac{k}{h^2} (u_{i-1}^n - 2u_i^n + u_{i+1}^n) + k f_i^{n+1} \right| , d_i \right] \\ \qquad\qquad (\text{pour } i = 1, \ldots N) \ , \\ u_N^{n+1} = \underline{\min} \left[\left| u_N^n + \frac{k}{h^2} (u_{N-1}^n - 2u_N^n) + k f_N^{n+1} \right| , d_N \right] . \end{array}\right.$$

Dans le tableau qui suit (Tableau 16) sont figurés les principaux résultats concernant le schéma explicite.

Le terme "explosion" signifie que la valeur de l'ECART entre la solution exacte et la solution approchée tend vers l'infini lorsque t augmente.

Figure 27



Représentation de la solution du problème de la torsion en évolution : solution exacte .

Tableau 16

S C H E M A E X P L I C I T E

Caractéristiques		Valeurs de $t \in [0, 1]$ $\frac{1}{\sqrt{h}}$ $ECART_2$ (solution exacte - solutions trouvées)											Temps de résolution
h	k	t=k	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.	
$\frac{1}{10}$	$\frac{1}{1000}$	0.	.0033	.0021	.0048	.0018	.0065	.0010	.0108	.0003	.0228	0.	540
	$\frac{1}{600}$	0.	.0035	.0023	.0051	.0020	.0069	.0011	.0110	.0004	.0232	0.	244
	$\frac{1}{500}$	0.	.0036	.0024	.0052	.0021	.0070	.0012	.0112	.0005	.0235	0.	207
	$\frac{1}{400}$	0.	.0038	.0026	.0055	.0023	.0073	.0013	.0114	.0006	.0239	0.	170
	$\frac{1}{250}$	.0001	.0042	.0033	.0062	.0030	.0080	.0018	.0123	.0010	.0250	0.	128
	$\frac{1}{200}$	.0001	.0046	.0038	.0067	.0034	.0085	.0022	.0128	.0012	.0250	0.	115
	$\frac{1}{150}$	.0003	.0058	.9162	28.9	1004.	35381.	10^6	10^7	10^9	10^{10}	10^{12}	101
explosion	$\frac{1}{100}$	.0006	.1126	135.	84054	5.10^7	10^{10}	10^{13}	10^{16}	10^{18}	10^{21}	10^{24}	95
$\frac{1}{20}$	$\frac{1}{1000}$	0.	.0010	.0013	.0014	.0014	.0012	.0010	.0007	.0004	.0003	0.	677
	$\frac{1}{800}$	0.	.0011	.0015	.0016	.0015	.0014	.0011	.0008	.0004	.0003	0.	580
	$\frac{1}{700}$	0.	.5630	3500.	--	--	--	--	--	--	--	--	--
explosion	$\frac{1}{100}$	.0008	3632	--	--	--	--	--	--	--	--	--	--

c) Etude du schéma implicite en équation

Le schéma s'écrit

$$\left|\begin{array}{l} u_1^{n+\frac{1}{2}} - \frac{k}{h^2}\,(u_2^{n+\frac{1}{2}} - 2.u_1^{n+\frac{1}{2}}) = u_1^n + k\,f_1^{n+1}\,, \\ u_i^{n+\frac{1}{2}} - \frac{k}{h^2}\,(u_{i+1}^{n+\frac{1}{2}} - 2.u_i^{n+\frac{1}{2}} + u_{i-1}^{n+\frac{1}{2}}) = u_i^n + k\,f_i^{n+1}\,, \\ u_N^{n+\frac{1}{2}} - \frac{k}{h^2}\,(-2.u_N^{n+\frac{1}{2}} + u_{N-1}^{n+\frac{1}{2}}) = u_N^n + k\,f_N^{n+1}\,, \\ u_i^{n+1} = \min\left[\,\left|u_i^{n+\frac{1}{2}}\right|\,,\ d_i\right]\,,\quad i = 1,\ \ldots,\ N\,. \end{array}\right.$$

Ce schéma a été résolu par les méthodes de

- Gauss-Seidel,
- gradient conjugué,

utilisées à chaque itération en temps.

Le test d'arrêt commun dans la recherche de la solution à l'instant $t = nk$ est

$$\sqrt{\sum_{i=1}^{N} \left|J'_{i,h}(v^n)\right|^2} \leq 10^{-6}$$

avec

$$\left|\begin{array}{l} J'_{1,h}(v^n) = (1+2\,\frac{k}{h^2})\,v_1^n - \frac{k}{h^2}\,v_2^n - u_1^{n-1} - k\,f_1^n\,, \\ J'_{i,h}(v^n) = (1+2\frac{k}{h^2})\,v_i^n - \frac{k}{h^2}\,v_{i-1}^n - \frac{k}{h^2}\,v_{i+1}^n - u_i^{n-1} - kf_i^n\,, \\ J'_{N,h}(v^n) = (1+2\frac{k}{h^2})\,v_N^n - \frac{k}{h^2}\,v_{N-1}^n - u_N^{n-1} - kf_i^n\,, \end{array}\right.$$

où u^{n-1} est la solution trouvée à l'instant $t = (n-1)k$.

Les résultats concernant la résolution de ce schéma par la méthode de Gauss-Seidel sont consignés dans le Tableau 18; une étude sur la variation du paramètre $\omega_{opt.théorique}$ calculé par la méthode itérative décrite dans ce même chapitre au § 1.6,f) est donné dans le Tableau 17; une représentation

graphique de cette variation est donnée à la Figure 28.

Les résultats concernant la méthode du gradient conjugué sont donnés dans le Tableau 18. Les valeurs u_i^n, i = 1,...,N; n = 1,..., NT obtenues par les deux méthodes sont en général les mêmes à 10^{-4} près.

Tableau 17

Variation du paramètre optimal de la méthode de Gauss-Seidel en fonction de h et k .

h	k	$\omega_{opt.\ théorique}$	Temps de calcul
$\frac{1}{10}$	$\frac{1}{600}$	1.01458	54
	$\frac{1}{400}$	1.02654	52
	$\frac{1}{200}$	1.06403	53
	$\frac{1}{150}$	1.08732	57
	$\frac{1}{100}$	1.12786	73
	$\frac{1}{50}$	1.21296	53
	$\frac{1}{10}$	1.40481	56
	$\frac{1}{5}$	1.45676	55
	$\frac{1}{2}$	1.49633	58
$\frac{1}{20}$	$\frac{1}{800}$	1.06990	91
	$\frac{1}{100}$	1.35260	108
	$\frac{1}{10}$	1.63948	112
	$\frac{1}{5}$	1.67912	113
$\frac{1}{30}$	$\frac{1}{100}$	1.49955	109
	$\frac{1}{10}$	1.74324	137

Figure 28

Variation du paramètre optimal $\omega_{opt.théor.}$ utilisé dans la méthode de Gauss-Seidel appliquée aux problèmes paraboliques sur $H_0^1(0,1)$ en fonction de h,k.

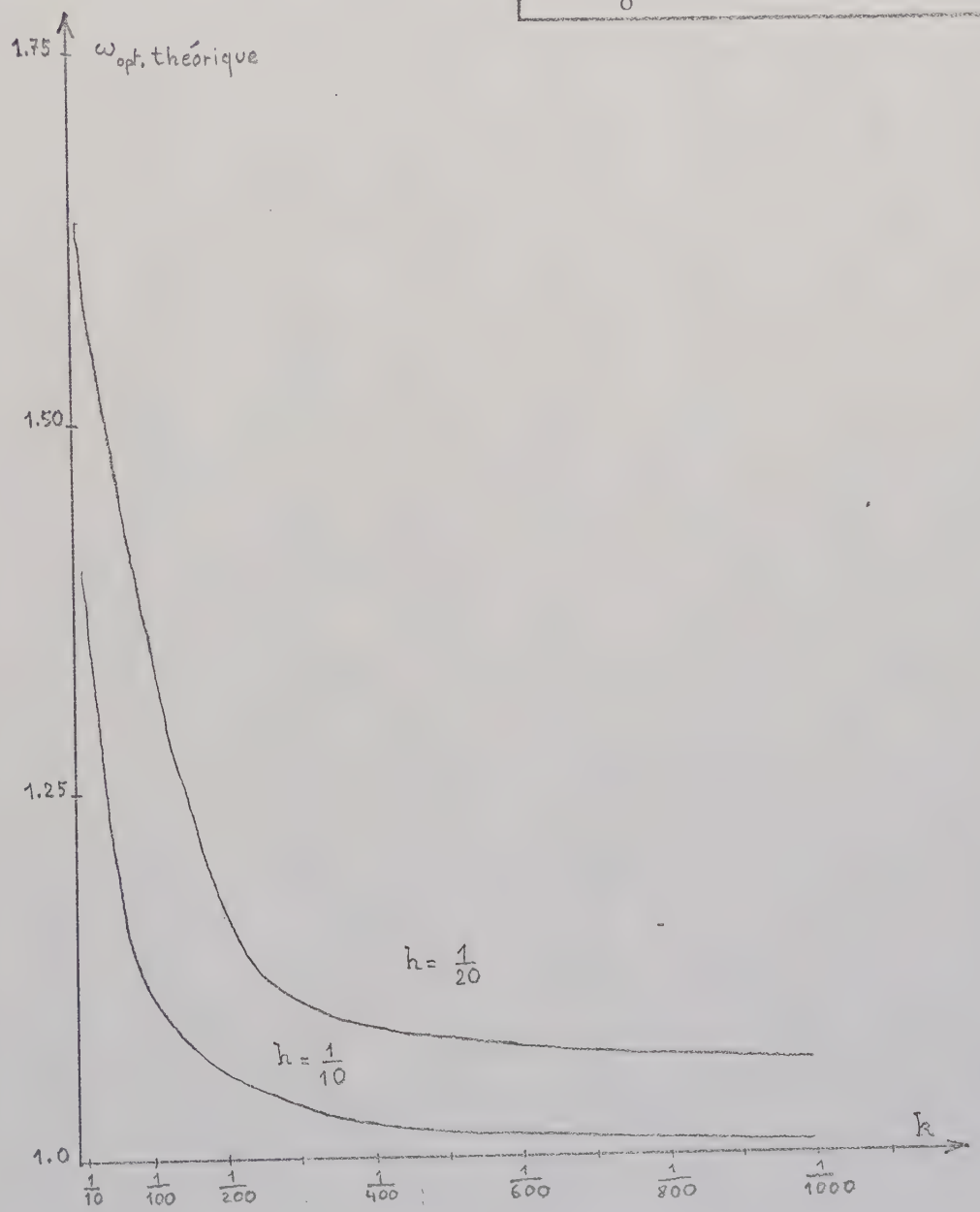


Tableau 18

SCHEMA IMPLICITE EN EQUATION { - Méthode du gradient conjugué.
- Méthode de Gauss-Seidel avec $\omega_{op.théorique}$.

Caractéristiques		$\frac{1}{\sqrt{h}}$ ECART$_2$ (solution exacte - solution trouvée)										Temps de résolution
Valeurs de $t \in [0, 1.] \rightarrow$	k	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.	
$h = \frac{1}{10}$	$k = \frac{1}{1000}$	2 / 0. / 3	2 / .0030 / 3	5 / .0085 / 5	4 / .0279 / 5	4 / .0252 / 5	4 / .0267 / 5	3 / .0182 / 5	3 / .0211 / 5	5 / .0191 / 5	3 / .0250 / 5	3 / 0. / 5	1302 / 1103
	$k = \frac{1}{600}$	2 / 0. / 4	2 / .0030 / 4	6 / .0134 / 6	5 / .0488 / 6	6 / .0417 / 6	5 / .0477 / 6	5 / .0417 / 6	5 / .0324 / 5	6 / .0427 / 5	4 / .0250 / 5	4 / 0. / 5	958 / 740
	$k = \frac{1}{400}$	2 / 0. / 5	2 / .0031 / 5	6 / .0205 / 7	5 / .0752 / 6	4 / .0523 / 6	4 / .0792 / 6	4 / .0533 / 6	5 / .0531 / 6	4 / .0500 / 6	4 / .0250 / 6	4 / 0. / 6	715 / 567
	$k = \frac{1}{150}$	3 / 0. / 8	2 / .0031 / 8	4 / .0382 / 10	8 / .1225 / 9	6 / .1148 / 10	6 / .1121 / 9	6 / .0871 / 9	6 / .0759 / 7	4 / .0500 / 7	3 / .0250 / 7	4 / 0. / 7	360 / 302
	$k = \frac{1}{100}$	3 / 0. / 8	3 / .0031 / 8	5 / .0842 / 10	6 / .1648 / 10	5 / .1592 / 10	6 / .1404 / 10	6 / .1061 / 10	6 / .0759 / 10	8 / .0500 / 10	6 / .0250 / 10	5 / 0. / 9	232 / 242
	$k = \frac{1}{50}$	3 / 0. / 11	4 / .0032 / 11	4 / .1489 / 13	4 / .2020 / 12	4 / .1787 / 11	4 / .1404 / 11	5 / .1061 / 12	5 / .0759 / 12	5 / .0500 / 12	5 / .0250 / 12	6 / 0. / 12	138 / 169
$h = \frac{1}{20}$	$k = \frac{1}{800}$	3 / 0. / 6	4 / .0089 / 9	4 / .0355 / 10	4 / .0442 / 10	3 / .0452 / 10	4 / .0414 / 10	3 / .0409 / 10	4 / .0380 / 10	4 / .0340 / 9	5 / .0250 / 9	5 / 0. / 8	2436 / 2781

Commentaires Tableau 18: Dans les triangles supérieurs est figuré le nombre d'itérations requises par la méthode du gradient conjugué pour obtenir la solution à l'instant $t = nk$ correspondant. Dans les triangles inférieurs est figuré le nombre d'itérations requises par la méthode de Gauss-Seidel utilisée avec son paramètre optimal ($\omega_{\text{opt.théorique}}$: cf. Tableau 17) pour obtenir la solution à l'instant $t = nk$ correspondant.

Dans la dernière colonne est figuré, dans les triangles supérieurs, le temps total de résolution par la méthode du gradient conjugué, dans le triangle inférieur, le temps total de résolution par la méthode de Gauss-Seidel.

d) Etude du schéma implicite en projection

Le schéma s'écrit

$$\left|\begin{array}{l} \sum_{i=1}^{N} (u_i^{n+1} - u_i^n - \frac{k}{h^2}(u_{i+1}^{n+1} - 2.u_i^{n+1} + u_{i-1}^{n+1}) - kf_i^{n+1}, v_i - u_i^{n+1}) \geq 0 \\ \quad \forall\, v = (v_1, \ldots, v_N) \quad \text{avec } |v_i| \leq d_i,\ i = 1, \ldots, N; \\ \quad u^{n+1} = (u_1^{n+1}, \ldots, u_N^{n+1}) \text{ tel que } |u_i^{n+1}| \leq d_i,\ i = 1, \ldots, N \\ \quad (\text{on pose } u_o^{n+1} = u_{N+1}^{n+1} = 0 \quad \text{et } u^o \text{ est donné}). \end{array}\right.$$

Ce schéma peut être résolu en cherchant à minimiser sur $K = \{ v \mid |v_i| \leq d_i \}$ la fonction $J_h(v)$ qui a pour gradient $J_h'(v) = (J_1'(v), \ldots, J_N'(v))$ avec

$$J_i'(v) = v_i - u_i^n - \frac{k}{h^2}(v_{i+1} - 2.v_i + v_{i-1}) - kf_i^{n+1}$$

Le schéma a été résolu par la méthode de Gauss-Seidel en projection avec paramètre optimal et la méthode S.N.

Pour les deux méthodes le critère d'arrêt dans la recherche de la solution à l'instant $t = nk$ est

$$\sqrt{\sum_{i=1}^{N} |\tilde{J}_i'(v^n)|^2} \leq 10^{-6} ,$$

avec

$$\tilde{J}_i'(v^n) = \begin{cases} J_i'(v^n) & \text{si } |v_i^n| < d_i \text{ ,} \\ J_i'(v^n) & \text{si } |v_i^n| = d_i \text{ et } J_i'(v^n).\text{signe } (v_i^n) \geq 0 \text{ ,} \\ 0 & \text{si } v_i^n = d_i \text{ et } J_i'(v^n).\text{signe } (v_i^n) < 0 \text{ .} \end{cases}$$

Les résultats concernant la résolution de ce schéma par la méthode S.N sont donnés dans le Tableau 19 (on obtient sensiblement les mêmes résultats pour $ECART_2$ par la méthode de Gauss-Seidel en projection).

Une comparaison sur l'utilité des méthodes de Gauss-Seidel en projection et S.N est établie dans le Tableau 20.

Tableau 19

SCHEMA IMPLICITE EN PROJECTION

		$\frac{1}{\sqrt{h}}$ ECART$_2$ (solution exacte - solution trouvée) pour différentes valeurs de t									
h	k	t = k	= 0.1	= 0.2	= 0.3	= 0.4	= 0.5	= 0.6	= 0.7	= 0.8	= 0.9	= 1.
$\frac{1}{10}$	$\frac{1}{100}$	10^{-5}	.0031	.0013	.0044	.0012	.0063	.0006	.0100	.0001	.022	0.
	$\frac{1}{50}$	10^{-4}	.0032	.0012	.0046	.0011	.0065	.0006	.0106	.0002	.022	0.
	$\frac{1}{10}$	→	.0038	.0004	.0055	.0007	.0079	.0006	.0121	.0003	.024	0.
	$\frac{1}{5}$	→	/	.0011	/	.0009	/	.0004	/	10^{-5}	/	0.
	$\frac{1}{2}$	→	/	/	/	/	.0081	/	/	/	/	0.
$\frac{1}{20}$	$\frac{1}{100}$	10^{-5}	.0003	.0005	.0005	.0006	.0005	.0004	.0003	.0001	10^{-5}	0.
	$\frac{1}{10}$	→	.0008	.0010	.0010	.0008	.0006	.0004	.0002	.0001	10^{-6}	0.
	$\frac{1}{5}$	→	/	.0017	/	.0014	/	.0007	/	.0001	/	0.
$\frac{1}{30}$	$\frac{1}{100}$	10^{-5}	.0005	.0002	.0007	.0003	.0010	.0003	.0018	.0001	.0043	0.
	$\frac{1}{10}$	→	.0005	.0011	.0005	.0008	.0009	.0003	.0021	.0001	.0046	0.

Commentaires: Les barres obliques signifient que les valeurs correspondantes n'ont "malheureusement" pas été demandées à l'ordinateur. La flèche signifie que k est plus grand ou égal à 0.1.

Tableau 20

SCHEMA IMPLICITE EN PROJECTION
Comparaison des méthodes de Gauss-Seidel en projection et S.N.

Nombre d'itérations aux instants $t \in [0,T]$ ci-dessous :

h	k	Méthodes	0.1	0.2	0.5	0.8	1.	Temps
1/10	1/100	G.S.P.	8	7	6	5	1	201
		S.N	1 + 5	2 + 4	2 + 3	2 + 0	1 + 0	244
	1/50	G.S.P.	10	8	8	7	1	151
		S.N	1 + 5	1 + 5	2 + 3	2 + 0	1 + 0	170
	1/10	G.S.P.	19	15	16	15	1	102
		S.N	3 + 6	3 + 6	4 + 3	4 + 0	2 + 0	112
	1/5	G.S.P.	-/-	20	19	17	1	74
		S.N	-/-	3 + 8	4 + 2	4 + 0	2 + 0	94
	1/2	G.S.P.	-/-	-/-	23	-/-	1	65
		S.N	-/-	-/-	5 + 4	-/-	6 + 0	87
1/20	1/100	G.S.P.	16	15	11	10	1	624
		S.N	2 + 9	2 + 8	2 + 5	2 + 2	1 + 0	584
	1/10	G.S.P.	34	33	30	31	1	212
		S.N	3 + 11	4 + 8	4 + 12	4 + 2	2 + 0	151
	1/5	G.S.P.	-/-	42	40	39	1	144
		S.N	-/-	5 + 10	6 + 5	6 + 2	4 + 0	115
1/30	1/100	G.S.P.	25	23	17	16	1	1336
		S.N	4 + 14	2 + 12	4 + 8	2 + 3	1 + 0	1159
	1/50	G.S.P.	53	51	48	47	1	372

Commentaires Tableau 20 : La méthode de Gauss-Seidel en projection est utilisée avec le paramètre $\omega_{opt.théorique}$.
Pour la méthode S.N. nous donnons deux chiffres: le premier correspond au nombre d'itérations de gradient ou de gradient projeté, le second donne le nombre d'itérations utilisant le gradient conjugué.

e) Conclusions

En ce qui concerne le problème de l'elasto-plasticité en évolution nous avons les conclusions suivantes :

1- Schéma explicite
- programmation très simple ,
- importance d'un bon choix de k et h : $(\frac{k}{h^2} \leqslant \text{cste} \sim \frac{1}{5})$,
- explosion si k est trop grand ,
- donc, en absence d'évaluation de la constante, cette méthode impose un "tâtonnement numérique".

2- Schéma implicite en équation
- programmation assez simple (quand la "projection" n'offre pas de difficulté),
- pour un même choix de k et h, le schéma semble être moins précis que le schéma explicite ,
- pour un même choix de k et h, le schéma est deux fois moins rapide que le précédent ,
- le danger d'explosion semble réduit.

3- Schéma implicite en projection
- programmation plus délicate ,
- pas de danger d'explosion ,
- précision remarquable ,
- pour un même choix de k et h, le schéma est plus rapide que le précédent, et, à précision égale plus rapide que le schéma explicite.

4- Méthode de Gauss-Seidel en projection
- programmation très simple ,
- problème du calcul de ω_{opt} [1].

(1) qui en fait, dans le cas avec contraintes, n'a plus rien d'optima il faudrait prendre le ω_{opt} du problème de Dirichlet défini "à l'intérieur de la frontière libre".

5- Méthode S.N.

- programmation délicate ,
- un peu moins rapide que la précédente lorsqu'il y a très peu de variables (mais l'approximation est alors très grossière),
- la meilleure méthode lorsqu'il y a plus de 20 variables.

2.7. Un problème de frottement évolutif

Pour $t \in [0,T]$, on considère l'inéquation

$$(u', v-u) + a(u, v-u) + j(v) - j(u) \geq (f, v-u), \quad \forall v \in V, u \in \vartheta_o$$

avec

$$\Omega =]0,1[$$
$$T = 1$$
$$V = \left\{ v \mid v \in H^1(\Omega), \quad v\big|_{x=0} = 0 \right\}$$
$$\vartheta_o = \left\{ v \mid v \in W(0,T), \quad v(o) = u_o \text{ donné } \in V \right\}$$
$$a(u,v) = \int_\Omega \text{grad } u \text{ grad } v \, dx$$
$$j(v) = g \, |v|_{x=1}$$
$$g = 10 \left| t\left(\frac{3}{4} - t\right) \right|$$
$$f = 20x\left(\frac{3}{2} - x\right)\left(\frac{3}{4} - 2t\right) + 40t\left(\frac{3}{4} - t\right) .$$

Ce problème est encore équivalent à

$$u' - \Delta u = f \qquad \text{dans } \Omega ,$$
$$\left|\frac{\partial u}{\partial n}\right| \leq g, \qquad \text{en } x = 1 ,$$
$$u \frac{\partial u}{\partial n} + g|u| = 0, \quad \text{en } x = 1 .$$

Avec les données ci-dessus l'inéquation a comme solution exacte

$$u = 20x \left(\frac{3}{2} - x\right) t \left(\frac{3}{4} - t\right) .$$

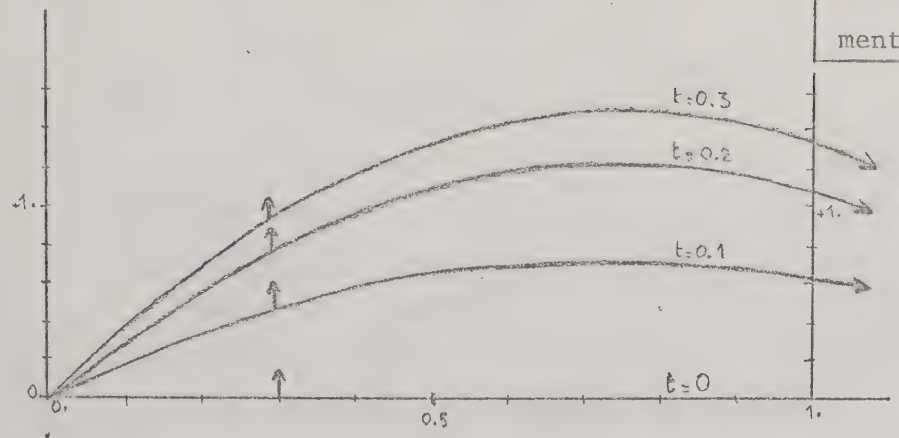
Cette solution est représentée sur la Figure 29 ci-après.

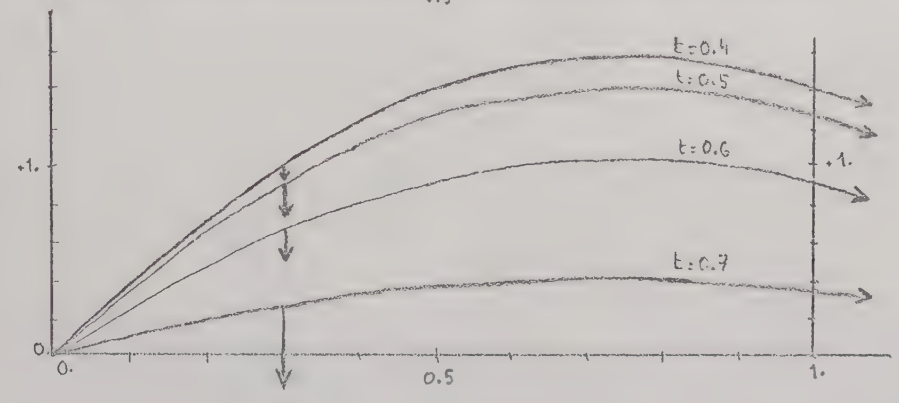
a) Discrétisation

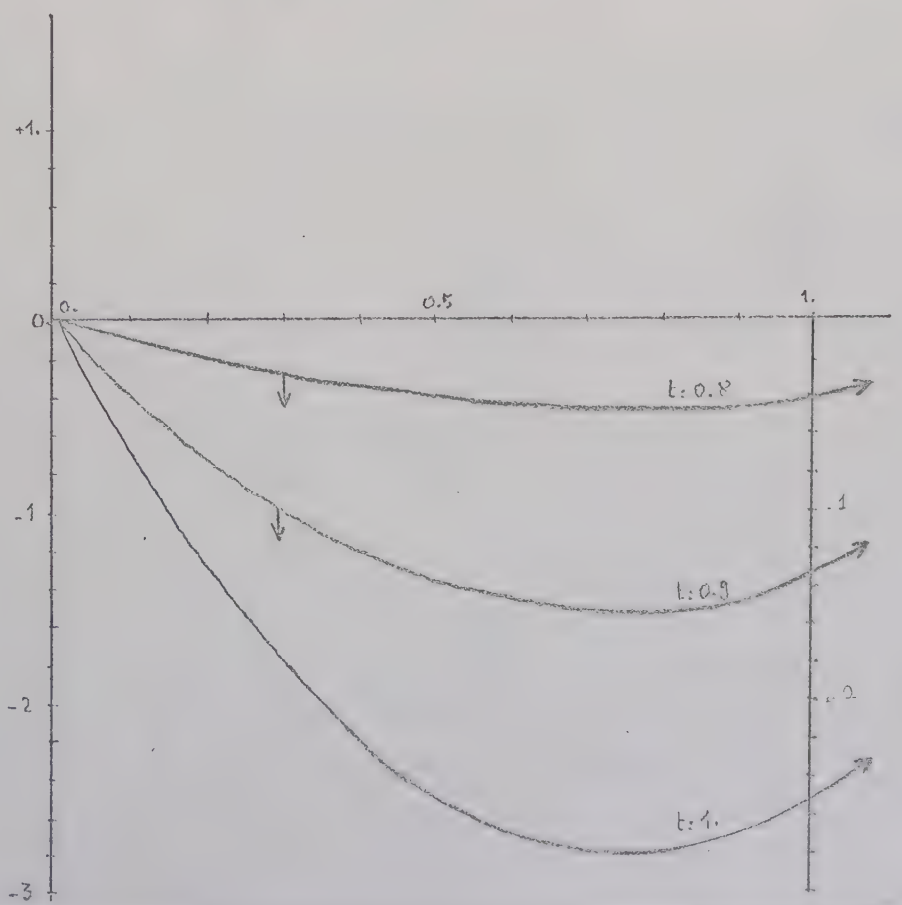
Pour la discrétisation en espace sur $]0,1[$ on cherche u_h de la forme $u_h = \sum_{i=1}^{N} u_i \, \theta_h^i$, ($V_h$ sera l'espace de ces fonctions

Figure 29

Représentation de la solution exacte du problème de frottement évolutif.







où θ_h^i est la fonction caractéristique de l'intervalle $\left[(i-\frac{1}{2})h,\ (i+\frac{1}{2})h\right[$ pour $i = 1, \ldots, N-1$ et θ_h^N est celle de $\left[1-\frac{h}{2},\ 1\right[$ avec N entier $= \frac{1}{h}$.

Nous allons indiquer rapidement comment l'on peut définir les schémas:

Obtention de schémas totalement implicites

Pour la commodité on se place d'abord dans le cas des approximations semi-discrètes (discrétisation en temps et non en espace). L'inéquation s'écrit

$$(I)\quad \left|\begin{array}{l} \left(\frac{u^{n+1}-u^n}{k},\ v-u^{n+1}\right) + \int_\Omega \text{grad } u^{n+1}.\text{grad}(v-u^{n+1})dx - \\ - \int_\Omega f^{n+1}(v-u^{n+1})dx + j^{n+1}(v) - j^{n+1}(u^{n+1}) \geqslant 0, \\ \forall\, v \in V,\ u^{n+1} \in V\ , \\ u^o = u_o \text{ donné.} \end{array}\right.$$

Il est aisé de voir que cette inéquation est équivalente au problème

$$(II)\quad \left|\begin{array}{l} \underline{\text{minimiser}}_{u^{n+1} \in V} \quad \frac{1}{2}\int_\Omega |u^{n+1}|^2 dx - \int_\Omega u^{n+1}u^n dx + \frac{k}{2}\int_\Omega |\text{grad } u^{n+1}| \\ \qquad - k\int_\Omega f^{n+1}u^{n+1}dx + kj^{n+1}(u^{n+1}) \end{array}\right.$$

ou encore, grâce à la forme particulière de j(v), à résoudre

$$(III)\quad \left|\begin{array}{l} \underline{\max}_{|p|\leqslant 1}\ \underline{\min}_{u^{n+1}\in V} \quad \frac{1}{2}\int_\Omega |u^{n+1}|^2 dx - \int_\Omega u^{n+1}u^n dx + \frac{k}{2}\int_\Omega |\text{grad } u^{n+1}| \\ \qquad - k\int_\Omega f^{n+1}u^{n+1}dx + k\,p.g^{n+1}u^{n+1}\ (x=1)\ . \end{array}\right.$$

On discrétise alors le problème en espace, soit (en divisant par h, en notant que la mesure du dernier intervalle est $\frac{h}{2}$, et en posant $u_o^{n+1} = 0$):

$$(IV)\quad \max_{|p|\le 1}\ \min_{v_h^n \in V_h} L_h^n(v_h^n, p) = \frac{1}{2}\sum_{i=1}^{N-1} \left| u_i^{n+1}\right|^2 + \frac{1}{4}\left|u_N^{n+1}\right|^2 -$$

$$- \left| \sum_{i=1}^{N-1} u_i^{n+1}\, u_i^n - \frac{1}{2}\, u_N^{n+1}\, u_N^n + \frac{k}{2h^2}\sum_{i=1}^{N} \left| u_i^{n+1} - u_{i-1}^{n+1}\right|^2 \right. ,-$$

$$- \left| \sum_{i=1}^{N-1} k f_i^{n+1}\, u_i^{n+1} - \frac{k}{2}\, f_N^{n+1}\, u_N^{n+1} + \frac{k}{h}\, p\, g^{n+1}\, u_N^{n+1} \right. .$$

Puis l'on écrit les conditions d'optimalité d'Euler: trouver $u_1^{n+1}, \ldots, u_N^{n+1}, p$ tels que

$$(V)\quad \frac{u_i^{n+1} - u_i^n}{k} - \frac{1}{h^2}\,(u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^n) - f_i^{n+1} = 0,\ i = 1, \ldots N-1$$

$$(VI)\quad \frac{u_N^{n+1} - u_N^n}{k} + \frac{2}{h^2}\,(u_N^{n+1} - u_{N-1}^{n+1}) - f_N^{n+1} + \frac{2}{h}\, p\, g^{n+1} = 0$$

$$(VII)\quad \left(\frac{k}{h}\, g^{n+1}\, u_N^{n+1},\ q - p\right) \ge 0, \qquad \forall q,\ |q| \le 1, \text{ et } |p| \le 1.$$

De la formulation (IV), on peut se référer à une méthode duale semi-implicite[1] pour définir l'algorithme suivant, qui pour u_h^n connu permet de déterminer u_h^{n+1} comme suit:

- prendre p° quelconque, par exemple $p^\circ = \text{signe}(u_N^n) \times 1.$, et un pas $\rho > 0$, puis
- résoudre (V), (VI), pour $p = p^\circ$; ce qui donne en particulier $u_N^{n+1}(o)$,
- faire $p^1 = P\left[p^\circ + \rho\, g^{n+1}\, u_N^{n+1}(0)\right]$, P = projection sur $[0,1]$
- résoudre (V), (VI) pour $p = p^1$; ce qui donne en particulier $u_N^{n+1}(1)$,
- faire $p^2 = P\left[p^1 + \rho\, g^{n+1}\, u_N^{n+1}(1)\right]$ etc...

On peut définir de multiples variantes de ces schémas; par exemple on peut remplacer dans (IV)

$$\frac{1}{h^2}\left| u_N^{n+1} - u_{N-1}^{n+1}\right|^2$$

---

(1) Ou en notant que (VII) est équivalent à $(p - p + \rho\, g^{n+1}\, u_N^{n+1},\ p - q) \ge 0$ $\forall q,\ |q| \le 1,\ |p| \le 1$

par le carré de

$$\frac{\partial}{\partial n_h}(u_h)_{x=1} = \frac{1}{2h}\left[3u_N - 4u_{N-1} + u_{N-2}\right]$$

où cette quantité n'est autre qu'une approximation "quadratique" de la dérivée (dérivée d'un polynôme du second degré passant par trois points[1]).

De même dans (V), (VI) on peut remplacer, pour $n \geq 1$,

$$\frac{u_i^{n+1} - u_i^n}{k} \quad \text{par} \quad \frac{1}{2k}\left[3u_i^{n+1} - 4u_i^n + u_i^{n-1}\right].$$

Obtention du schéma totalement explicite

On remplace[2] (V), (VI), (VII) par

$$\text{(VIII)} \quad \frac{u_i^{n+1} - u_i^n}{k} - \frac{1}{h^2}\left(u_{i+1}^n - 2u_i^n + u_{i-1}^n\right) - f_i^{n+1} = 0, \; i = 1, \ldots, N-1,$$

$$\text{(IX)} \quad \frac{u_N^{n+1} - u_N^n}{k} + \frac{2}{h^2}\left(u_N^n - u_{N-1}^n\right) - f_N^{n+1} + \frac{2}{h}\, p\, g^{n+1} = 0 ,$$

$$\text{(X)} \quad p = \text{signe}\,(u_N^n) \times 1.$$

b) Résultats numériques pour les schémas totalement explicites et totalement implicites

Afin d'étudier la valeur de ces schémas et de l'approximation nous définissons les quantités suivantes:

- la dérivée normale de la solution exacte en x = 1 et à t = nk,

$$\frac{\partial u^n}{\partial n}\text{ex.}\Big|_{x=1} = -10.t\left(\frac{3}{4} - t\right), \quad t = nk ;$$

- la dérivée normale discrète de la solution exacte en x = 1 et à t = nk,

$$\nabla_n u^n_{\text{ex.}}\Big|_{x=1} = \frac{u(1,nk) - u(1-h,nk)}{h} ,$$

---

(1) De façon générale la dérivée en un point x d'un polynome du second degré dont on connaît les valeurs y_o, y_1, y_2 en 3 points x, x-h, x-2h est $\frac{1}{2h}\left[3y_o - 4y_1 + y_2\right]$.

(2) Mais nous n'avons pas donné de justification théorique à ce schéma.

- la dérivée normale discrète de la solution approchée en $x = 1$ et à $t = nk$,

$$\nabla_h u_h^n \Big|_{x=1} = \frac{u_N^n - u_{N-1}^n}{k} ,$$

- l'$ECART_2$ entre la solution exacte et la solution trouvée à $t = nk$

$$ECART_2^{t=nk} = \sqrt{\sum_{i=1}^{N} (u(ih, nk) - u_i^n)^2 \, h} .$$

Les deux schémas ont été résolus sur le BULL-GE-265 de la Société General Time Sharing. Les temps sont donnés en 1/6e de seconde, compilation comprise.

Nous donnons dans les tableaux suivants (Tableaux 21 et 22) les résultats concernant le schéma explicite (VIII, IX, X) et le schéma implicite (V, VI, VII).

Tableau 21

FROTTEMENT EVOLUTIF : SCHEMA EXPLICITE

h	k	ECART$_2$ t = k	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	Temps
1/10	1/100	.004	18.	10^6	10^{11}	10^{16}	10^{21}	10^{35}	10^{30}	10^{35} ... EXPLOSION			--
	1/500	.0001	.0015	.0012	.0005	.0014	.0033	.0056	.0083	.0111	.0142	.0174	145
	1/1000	.000045	.00078	.00064	.00025	.00070	.00167	.00284	.00415	.00558	.00710	.00869	247
	1/2000	.000011	.00039	.00032	.00012	.00035	.00083	.00142	.00207	.00276	.00353	.00434	463
1/20	1/100	.0065	10^8	10^{20}	10^{32}	... EXPLOSION							--
	1/500	.0026	10^{13}	10^{30}	... EXPLOSION								--
	1/1000	.000066	.00077	.00064	.00024	.00068	.00164	.00280	.00410	.00552	.00702	.00859	427
	1/2000	.000016	.00038	.00032	.00012	.00034	.00082	.00140	.00205	.00276	.00351	.00429	802
1/30	1/100	... EXPLOSION											--
	1/500	... EXPLOSION											--
	1/1000	.000081	10^{36}	... EXPLOSION									--
	1/2000	.000020	.00038	.00032	.00011	.00033	.00082	.00139	.00204	.00274	.00349	.00427	1548

Tableau 22

FROTTEMENT EVOLUTIF : SCHEMA IMPLICITE(1)

h	k	ECART$_2$ t = k	0.1	0.2	0.3	0.4	0.5	0.6	0.75(2)	0.8	0.9	1.0	Temps
1/10	1/100	90.10^{-5}	81.10^{-4}	14.10^{-3}	19.10^{-3}	23.10^{-3}	26.10^{-3}	28.10^{-3}	31.10^{-3}	32.10^{-3}	33.10^{-3}	34.10^{-3}	210
	1/500	36.10^{-6}	16.10^{-4}	29.10^{-4}	39.10^{-4}	47.10^{-4}	53.10^{-4}	58.10^{-4}	63.10^{-4}	64.10^{-4}	67.10^{-4}	68.10^{-4}	506
	1/1000	91.10^{-7}	81.10^{-5}	14.10^{-4}	19.10^{-4}	23.10^{-4}	26.10^{-4}	29.10^{-4}	31.10^{-4}	32.10^{-4}	33.10^{-4}	34.10^{-4}	840
1/20	1/100	88.10^{-5}	79.10^{-4}	14.10^{-3}	10.19^{-3}	22.10^{-3}	25.10^{-3}	28.10^{-3}	30.10^{-3}	31.10^{-3}	32.10^{-3}	33.10^{-3}	586
	1/500	36.10^{-6}	16.10^{-4}	28.10^{-4}	38.10^{-4}	46.10^{-4}	52.10^{-4}	56.10^{-4}	61.10^{-4}	63.10^{-4}	65.10^{-4}	67.10^{-4}	1349
	1/1000	90.10^{-7}	80.10^{-5}	14.10^{-4}	19.10^{-4}	23.10^{-4}	26.10^{-4}	28.10^{-4}	30.10^{-4}	31.10^{-4}	32.10^{-4}	33.10^{-4}	2029
1/30	1/100	88.10^{-5}	78.10^{-4}	14.10^{-3}	18.10^{-3}	22.10^{-3}	25.10^{-3}	28.10^{-3}	30.10^{-3}	31.10^{-3}	32.10^{-3}	33.10^{-3}	1183
	1/500	36.10^{-6}	15.10^{-4}	28.10^{-4}	38.10^{-4}	45.10^{-4}	51.10^{-4}	56.10^{-4}	61.10^{-4}	62.10^{-4}	64.10^{-4}	66.10^{-4}	2786
	1/1000	90.10^{-7}	80.10^{-5}	14.10^{-4}	19.10^{-4}	23.10^{-4}	25.10^{-4}	28.10^{-4}	30.10^{-4}	31.10^{-4}	32.10^{-4}	33.10^{-4}	4333

Commentaires: (1) Résolu par la méthode de Gauss-Seidel avec utilisation de $\omega_{opt.expérimental}$ obtenu par tâtonnement. On a retenu, en fonction de h et k

Tableau 23

h	1/10			1/20			1/30		
k	1/100	1/500	1/1000	1/100	1/500	1/1000	1/100	1/500	1/1000
$\omega_{opt.exp.}$	1.4	1.1	1.1	1.45	1.15	1.1	1.5	1.25	1.12
itérations	11	6	5	20	9	7	28	13	10

La dernière ligne de ce tableau indique le nombre moyen d'itérations nécessaire pour obtenir $\|\mathcal{L}^n_h(v^n_h, p)\|_2 \leq 10^{-2}$.

Afin de préciser la valeur des schémas, nous donnons dans le tableau suivant (Tableau 24) quelques estimations d'erreur entre les valeurs de la solution exacte discrétisée et les valeurs des solutions approchées trouvées.

En fonction de h et k nous donnons

$$\nabla = \max_{n=1,\ldots,NT=\frac{1}{k}} \left| \nabla_h u^n_{ex.} - \nabla_h u^n_h \right|_{x=1} ,$$

$$\cup = \max_{\substack{n=1,\ldots,NT=\frac{1}{k} \\ i=1,\ldots,N=\frac{1}{h}}} \left| u(ih, nk) - u_h(ih, nk) \right| . \blacksquare$$

Sur la Figure 30 on a représenté en fonction de $t \in [0,1]$ les deux quantités

$$\frac{\partial u}{\partial n} \text{ ex.} \Big|_{x=1} \qquad \text{et} \qquad \nabla_h u_{ex.} \Big|_{x=1} .$$

respectivement dérivée exacte en x = 1 de la solution exacte et dérivée discrète(pas h) en x = 1 de la solution exacte. ■

Tableau 24

PRECISION RELATIVE DES SCHEMAS					
		SCH. EXPLICITE		SCH. IMPLICITE	
h	k	U	∇	U	∇
$\frac{1}{10}$	$\frac{1}{100}$	E X P L O S I O N		8.10^{-3}	2.10^{-4}
	$\frac{1}{500}$	2.10^{-3}	10^{-2}	2.10^{-3}	4.10^{-5}
	$\frac{1}{1000}$	$1.5\ 10^{-3}$	6.10^{-3}	10^{-3}	2.10^{-5}
	$\frac{1}{2000}$	10^{-3}	5.10^{-3}	NON TRAITE	
$\frac{1}{20}$	$\frac{1}{100}$	E X P L O S I O N		8.10^{-3}	10^{-4}
	$\frac{1}{500}$	E X P L O S I O N		2.10^{-3}	2.10^{-5}
	$\frac{1}{1000}$	3.10^{-3}	5.10^{-3}	10^{-3}	10^{-5}
	$\frac{1}{2000}$	2.10^{-3}	2.10^{-3}	NON TRAITE	
$\frac{1}{30}$	$\frac{1}{100}$	E X P L O S I O N		8.10^{-3}	5.10^{-5}
	$\frac{1}{500}$	E X P L O S I O N		2.10^{-3}	10^{-5}
	$\frac{1}{1000}$	E X P L O S I O N		10^{-3}	5.10^{-6}
	$\frac{1}{2000}$	10^{-3}	2.10^{-3}	NON TRAITE	

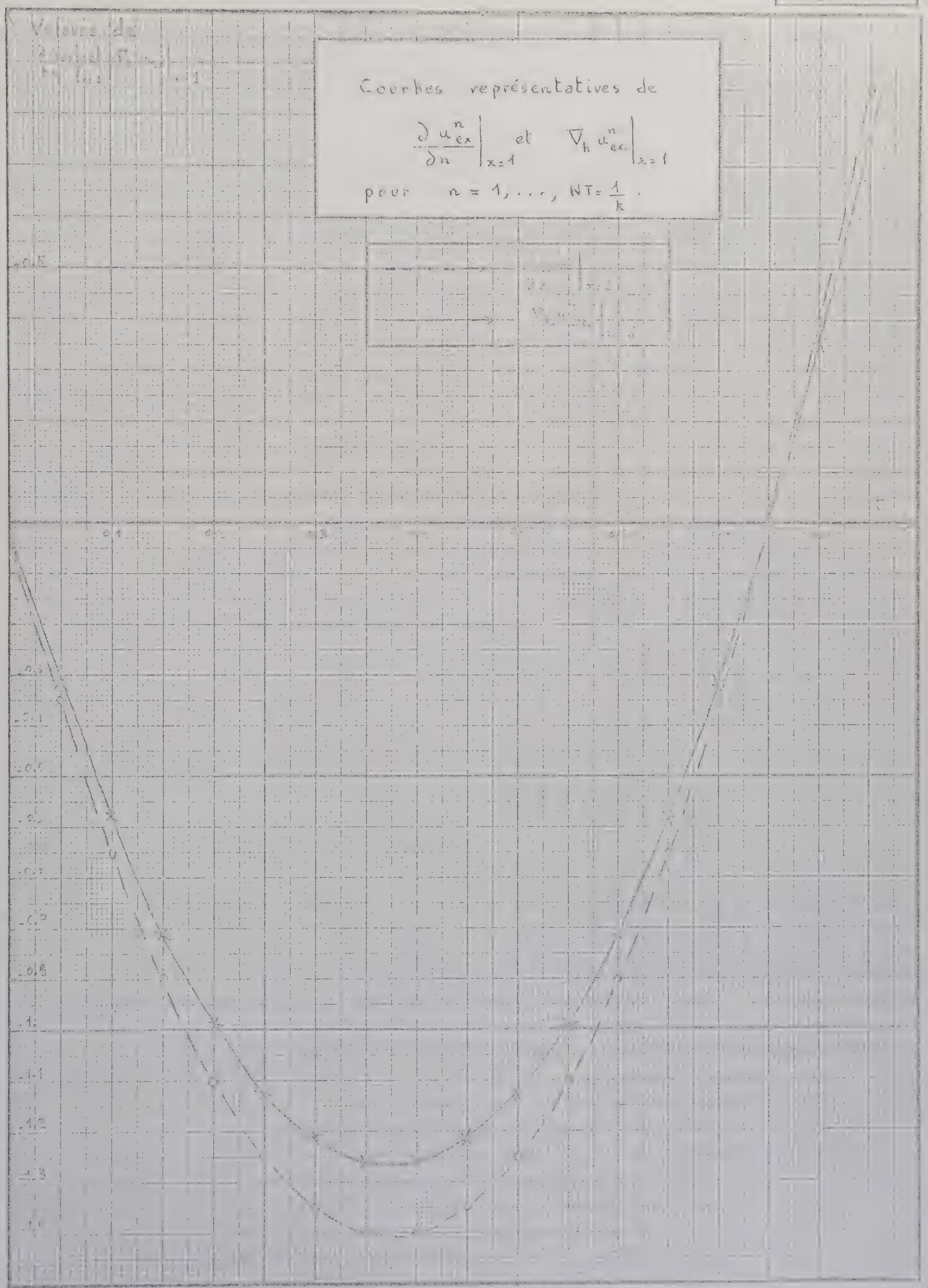
c) Conclusions

En ce qui concerne le problème de frottement évolutif considéré ici, nous avons

1. Schéma totalement explicite
   - danger d'explosion ,
   - la stabilité semble correspondre à $\frac{k}{h^2} \leqslant \frac{1}{2}$.

2. Schéma totalement implicite
   - précision remarquable pratiquement indépendante de h et k (mais la solution exacte est "quadratique") ,
   - à précision égale beaucoup moins rapide que le schéma explicite.

Figure 30


Courbes représentatives de
$\left.\frac{\partial u_{ex}^n}{\partial n}\right|_{x=1}$ et $\left.\nabla_h u_{ex}^n\right|_{x=1}$
pour $n = 1, \ldots, NT = \frac{1}{k}$.

2.8. Un problème parabolique à contraintes réparties

Pour t parcourant $[0,T]$ on se propose de résoudre l'inéquation

$$(u' - \Delta u - f,\ v - u) \geq 0, \quad \forall v \in K, \quad u \in \mathcal{K}$$

avec

$$K = \{ v \mid v \in H_o^1(\Omega),\ v \geq 0 \text{ p.p. sur } \Omega \},$$

$$\mathcal{K} = \{ v \mid v \in W(0,T),\ v(o) = u_o, \text{ donné},\ v \in K \text{ p.p. sur } Q = \Omega \times]0,T[\},$$

$$W(0,T) = \{ v \mid v \in L^2(0,T;V),\ v' \in L^2(0,T;V') \},$$

$$\Omega =]0,1[\times]0,1[.$$

On pose

$$a(t) = \frac{1}{2} + \frac{1}{4} \sin(2\pi t)$$

et l'on prend pour solution exacte

$$u = \begin{cases} -100\, x(x - a(t))^2\, y(1-y), & x \leq a(t), \\ 0, & x > a(t), \end{cases}$$

d'où la condition initiale

$$u_o = \begin{cases} -100\, x(x - \frac{1}{2})^2\, y(1-y), & x \leq \frac{1}{2}, \\ 0, & x > \frac{1}{2}. \end{cases}$$

Cette solution correspond à

$$f = \begin{cases} u' - \Delta u, & x \leq a(t), \\ \leq 0 \text{ quelconque}, & x > a(t). \end{cases}$$

On donne quelques courbes représentatives de la solution exacte sur les deux graphiques suivants (Figures 31 et 32).

Sur la Figure 33 sont représentées, en fonction du temps, les zones de Ω où $u = 0$ (convexe saturé).

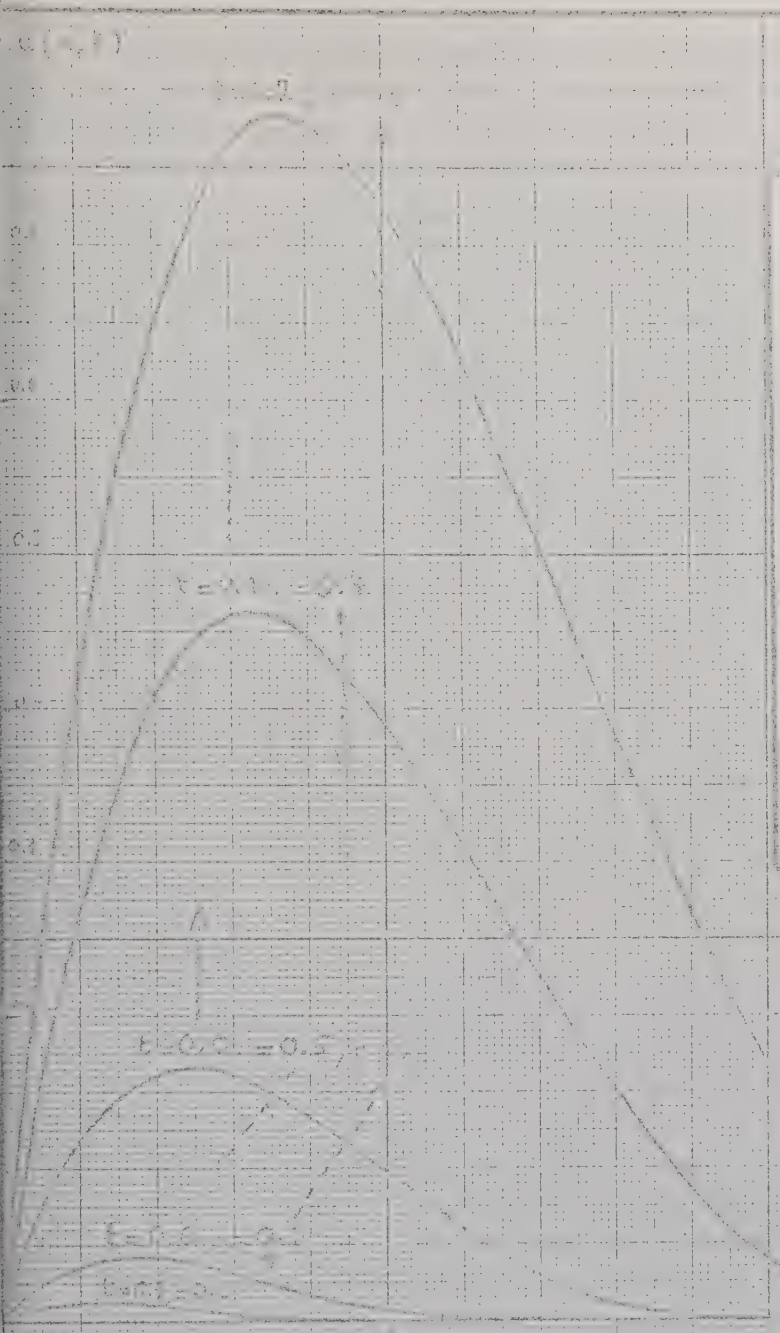
La discrétisation en espace est la discrétisation standard de $H_o^1(\Omega)$ avec

$$h = h_x = h_y = \frac{1}{20}, \text{ ou } \frac{1}{40}$$

(donc avec $N = \frac{1}{h} - 1$ on a N^2 variables)

$$u_{o,h} = \sum_{M \in \Omega_h} u_o(M)\ \theta_h^M,$$

Figure 31



Inéquation parabolique :

Coupes de la solution exacte pour $y = \frac{1}{2}$, aux instants $t = 0,\ 0.1,\ 0.2,\ \ldots,\ 1.$

$$(u' - \Delta u - f,\ v - u) \geq 0,$$
$$\forall v \in K,\ u \in K$$

$$u(0) = u_0 = \begin{cases} -100 \times \left(x - \frac{1}{2}\right) y (1-y) & \text{si } x \leq \frac{1}{2} \\ 0, & \text{si } x \geq \frac{1}{2} \end{cases}$$

$$K = \{ v \mid v \in H_0^1(\Omega),\ v \geq 0 \text{ sur } \Omega \}$$
$$\Omega =]0,1[\times]0,1[$$
$$t \in [0, 1]$$

← zône de saturation de K à t →

Figure 32

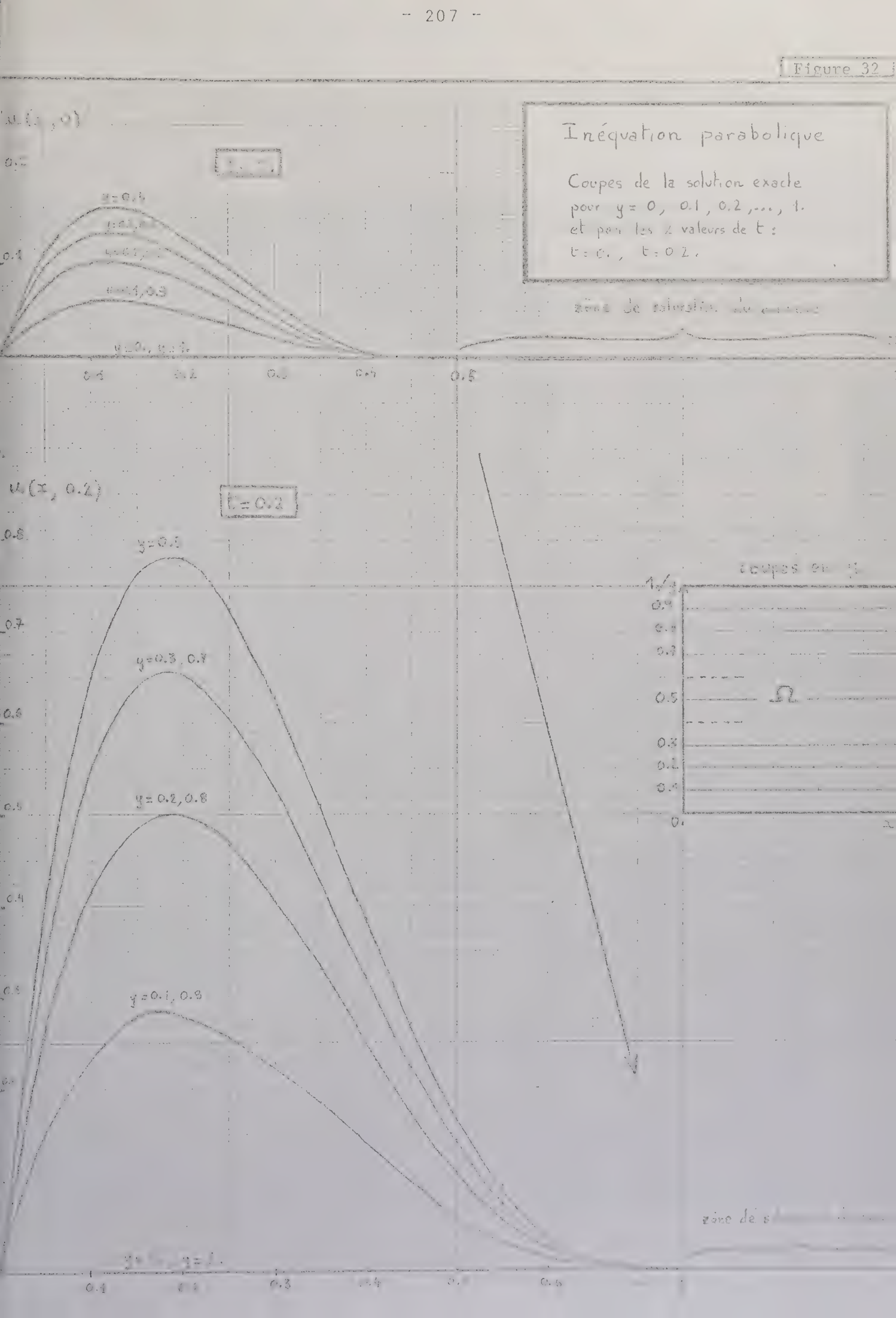
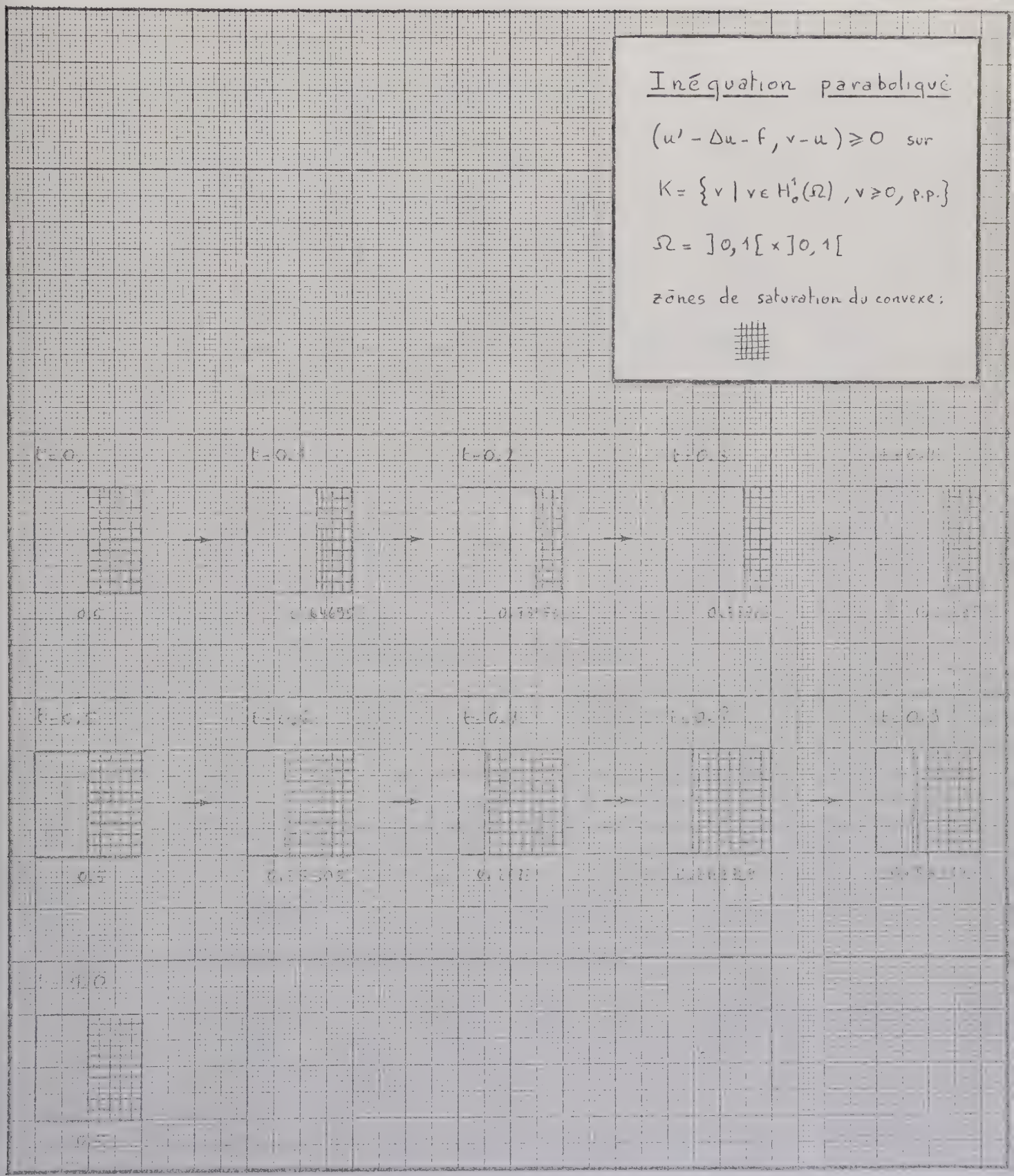

Inéquation parabolique
Coupes de la solution exacte
pour y = 0, 0.1, 0.2, ..., 1.
et pour les 2 valeurs de t :
t = 0., t = 0.2.
t = 0.2
y = 0.3
y = 0.2, 0.8
y = 0.1, 0.9
Ω

Figure 33


Inéquation parabolique
$(u' - \Delta u - f, v - u) \geq 0$ sur
$K = \{v \mid v \in H_0^1(\Omega), v \geq 0, \text{p.p.}\}$
$\Omega =]0,1[\times]0,1[$
zônes de saturation du convexe :

$$K_h = \left\{ v_h \mid v_h \in V_h \ , \quad v_h \geqslant 0 \text{ sur } \Omega_h \right\} .$$

Pour la discrétisation en temps nous prendrons

$$k = 0.5.10^{-3} \ , \ 10^{-3} \ , \ 2..10^{-3} \ , \ 5.10^{-3}$$

et $T = 1$.

Des essais ont été menés avec, pour $t \in [nk, (n+1)k[$;

$$f_{h,k}^{n} = \begin{cases} f_{h,k}^{n(-)} = \sum_{M \in \Omega_h} f(M; nk)\ \theta_h^M \\ f_{h,k}^{n(\frac{1}{2})} = \sum_{M \in \Omega_h} f(M; (n+\frac{1}{2})k)\ \theta_h^M \\ f_{h,k}^{n(+)} = \sum_{M \in \Omega_h} f(M; (n+1)k)\ \theta_h^M \end{cases}$$

Les schémas étudiés sont les suivants :

S1 - Schéma explicite

$$\left| \begin{array}{l} \dfrac{u_h^{n+\frac{1}{2}} - u_h^n}{k} - \Delta_h u_h^n - f_{h,k}^n = 0 \ , \\ u_h^{n+1} = P_{K_h}\left[u_h^{n+\frac{1}{2}} \right] \ , \ (1) , \\ u_h^o = u_{o,h} . \end{array} \right.$$

S2 - Schéma implicite en équation

$$\left| \begin{array}{l} \dfrac{u_h^{n+\frac{1}{2}} - u_h^n}{k} - \Delta_h u_h^{n+\frac{1}{2}} - f_{h,k}^n = 0 \ , \\ u_h^{n+1} = P_{K_h}\left[u_h^{n+\frac{1}{2}} \right] \ , \\ u_h^o = u_{o,h} . \end{array} \right.$$

S3 - Schéma semi-implicite en inéquation

$$\left| \begin{array}{l} \left(\dfrac{u_h^{n+1} - u_h^n}{k} - \Delta_h \dfrac{u_h^{n+1} + u_h^n}{2} - f_{h,k}^n \ , \ v_h - u_h^{n+1} \right) \geqslant 0, \forall v_h \in K_h, \ u_h^{n+1} \in K \\ u_h^o = u_{o,h} . \end{array} \right.$$

(1) P_{K_h} étant la projection définie par $(\ , \ )_h$.

§3 - Schéma totalement implicite

$$\left(\frac{u_h^{n+1} - u_h^n}{k} - \Delta_h u_h^{n+1} - f_{h,k}^n \, , \, v_h - u_h^{n+1} \right)_h \geqslant 0,$$

$$\forall v_h \in K_h \, , \; u_h^{n+1} \in K_h \, ,$$

$$u_h^o = u_{o,h} \, .$$

Excepté pour le schéma explicite, le calcul de u_h^{n+1} est obtenu par la méthode de Gauss-Seidel avec contraintes de bornes ($u_h^n \geqslant 0$), le paramètre de relaxation ω étant choisi de façon "optimale" par la formule (cf. VARGA[1])

$$\omega = \omega_{opt} = \frac{2}{1 + \left(1 - \left[\frac{(1 - \frac{\pi h^2}{2}) 4h^2}{4 + 2 \frac{h^2}{k^2}} \right]^2 \right)^{1/2}}$$

Le critère d'arrêt retenu dans le calcul de u_h^{n+1} est en général

$$\varepsilon_o = \sum_M \left| u_h^{n+1(m+1)} - u_h^{n+1(m)} \right| \leqslant 10^{-6} \, .$$

où $u_h^{n+1(m)}$ est le vecteur de l'itération m, et $u_h^{n+1(m+1)}$ celui de l'itération m+1. (Une itération correspondant à 1 balayage complet sur les coordonnées).

Avant de nous intéresser à la valeur relative des schémas nous allons d'abord regarder le comportement des solutions approchées en fonction du choix de $f_{h,k}^n = f_{h,k}^{n(-)}$, $f_{h,k}^{n(\frac{1}{2})}$, $f_{h,k}^{n(+)}$ du choix de $f \leqslant 0$ pour $x > a(t)$, et du test d'arrêt imposé à la méthode de Gauss-Seidel utilisée à chaque itération en temps.

La valeur relative des schémas sera mesurée à chaque instant $t = nk$ par la quantité

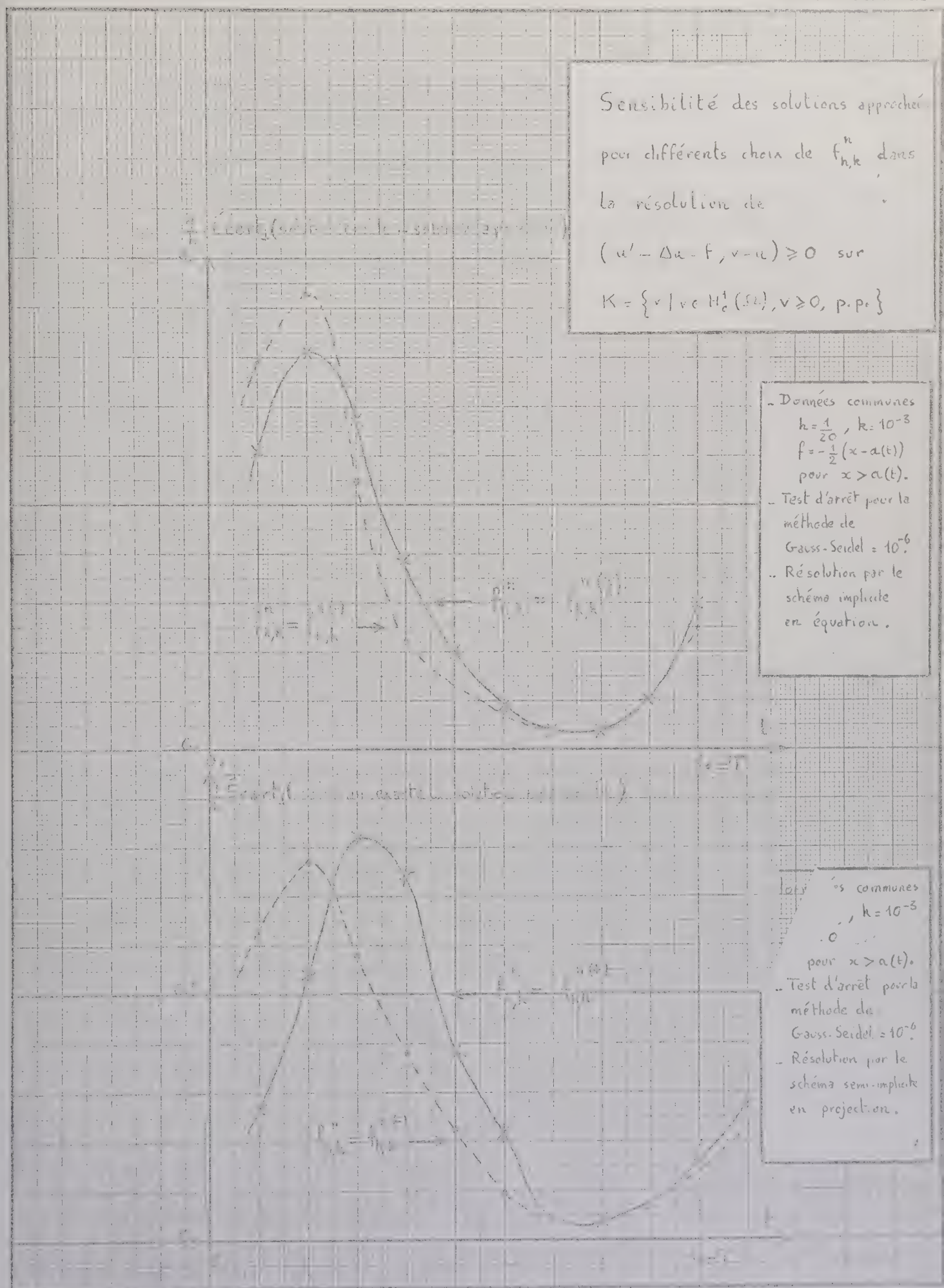
$$ECART_1 = \sum_{M \in \Omega_h} \left| u_h^n(M) - u(M; nk) \right| h$$

où u_h^n est la solution trouvée à $t = nk$ et $u(M; nk)$ est la solution exacte à $t = nk$, au point M.

a) Sensibilité des solutions approchées en fonction du choix de $f_{h,k}^n$

On verra sur le graphique suivant que le choix de $f_{h,k}^n$ semble être indifférent (Figure 34).

Figure 34


Sensibilité des solutions approchées pour différents choix de $f^n_{h,k}$ dans la résolution de
$(u' - \Delta u - f, v-u) \geq 0$ sur
$K = \{v \mid v \in H^1_0(\Omega), v \geq 0, \text{p.p.}\}$
- Données communes
$h = \frac{1}{20}$, $k = 10^{-3}$
$f = -\frac{1}{2}(x - a(t))$
pour $x > a(t)$.
- Test d'arrêt pour la méthode de Gauss-Seidel = 10^{-6}.
- Résolution par le schéma implicite en équation.
- Données communes
, $k = 10^{-3}$
pour $x > a(t)$.
- Test d'arrêt pour la méthode de Gauss-Seidel = 10^{-6}.
- Résolution par le schéma semi-implicite en projection.

b) Sensibilité des solutions approchées en fonction du choix de $f \leqslant 0$ pour $x \geqslant a(t)$

La solution exacte du problème initial ne dépendant pas du choix de $f \leqslant 0$ pour $x > a(t)$ il est intéressant de regarder s'il en est de même pour les solutions approchées. Le tableau suivant montre qu'en effet cela est vrai.
(Tableau 25).

Tableau 25

Sensibilité des solutions approchées pour différents choix de $f \leqslant 0$ pour $x \geqslant a(t)$.
On a donc fait varier $f \leqslant 0$ pour $x \geqslant a(t)$ avec les données communes suivantes :

$h = \frac{1}{20}$, $k = 10^{-3}$, $f^{n}_{h,k} = f^{n(+)}_{h,k}$; Test d'arrêt pour la méthode de Gauss-Seidel = 10^{-6}.
Résolutions par schéma semi-implicite en projection.

$f(x)$ pour $x > a(t)$	$\frac{1}{h}$ ECART$_1$ (solution exacte - solution approchée)									
	t = 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.
0	0.2788	0.5412	0.8191	0.7473	0.3830	0.1269	0.0438	0.0391	0.0799	0.1610
$-\frac{1}{2}(x-a(t))$	0.2777	0.5420	0.8230	0.7461	0.3829	0.1270	0.0438	0.0390	0.0790	0.1595
$-(x-a(t))$	0.2778	0.5423	0.8267	0.7462	0.3829	0.1271	0.0439	0.0391	0.0790	0.1594
$-4(x-a(t))$	0.2785	0.5434	0.8330	0.7467	0.3832	0.1275	0.0442	0.0394	0.0796	0.1598
$-100(x-a(t))$	0.3000	0.5614	0.8564	0.7629	0.3946	0.1375	0.0471	0.0403	0.0807	0.1613

c) Sensibilité des solutions approchées en fonction du test d'arrêt de la méthode de Gauss-Seidel.

On a regardé l'influence du test d'arrêt dans la méthode de Gauss-Seidel utilisée à chaque itération en temps pour la résolution du schéma implicite en équation.

On verra que les tests d'arrêt à 10^{-3} et 10^{-6} sont sensiblement équivalents. (Tableau 26). En fait, paradoxalement, la solution approchée est quelquefois meilleure pour le test d'arrêt à 10^{-3}.

Tableau 26

Résolution du schéma implicite en équation par la méthode de Gauss-Seidel pour deux tests d'arrêt différents. Les données communes sont $f = -\frac{1}{2}(x-a(t))$ pour $x > a(t)$ et $h = \frac{1}{20}$.

k	Test d'arrêt	$\frac{1}{h}$ ECART$_1$ (solution exacte - solution calculée)									
		t=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.
2.10^{-3}	10^{-3}	0.415	0.764	0.854	0.568	0.259	0.106	0.046	0.043	0.093	0.219
	10^{-6}	0.773	0.982	0.638	0.232	0.112	0.078	0.042	0.047	0.123	0.370
10^{-3}	10^{-3}	0.445	0.716	0.780	0.562	0.277	0.107	0.043	0.041	0.094	0.230
	10^{-6}	0.616	0.819	0.670	0.390	0.205	0.094	0.042	0.043	0.108	0.304

d) Précision du schéma explicite

Une première difficulté est le choix de k en fonction de h.

Ainsi, nous avons constaté que le schéma explicite "explosait" pour $h = \frac{1}{20}$ et $k = 10^{-3}$, dès les premières itérations en temps (résultats aberrants et dépassement de capacité avant $t = 100 \times k = 0.1$). Pour $h = \frac{1}{20}$, $k = 0.5.10^{-3}$ les résultats donnés par le schéma explicite sont corrects. Nous donnons dans le tableau suivant (Tableau 27) l'évolution de l'erreur entre la solution trouvée et la solution exacte.

Tableau 27

S C H E M A E X P L I C I T E											
$t = nk$	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55
$\frac{1}{n}$ $ECART_1$	0.19	0.28	0.39	0.51	0.67	0.78	0.81	0.74	0.56	0.38	0.22
$t = nk$	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.	1.05	1.10
$\frac{1}{n}$ $ECART_1$	0.12	0.07	0.04	0.039	0.037	0.05	0.08	0.11	0.16	0.20	0.28

Commentaires : $k = 0.5 \cdot 10^{-3}$, $h = \frac{1}{20}$, $f^n_{h,k} = f^{n(+)}_{h,k}$.

Temps total de résolution (jusqu'à t = 0.55) = 6 mn 225 | sur CII 10070
(jusqu'à t = 1.10) = 10 mn 529 | en Fortran H.

Pour $x \geq a(t)$ on a pris $f(x) = -\frac{1}{2}(x - a(t))$.

e) Précision du schéma implicite en équation

On donne dans le tableau suivant (Tableau 28) quelques résultats concernant la précision du schéma implicite en équation pour quelques valeurs de k et h.

On verra que pour un même pas de temps k, un pas d'espace h plus précis ne réduit paradoxalement pas l'erreur entre la solution approchée et la solution exacte; en fait, et ceci est en accord avec la théorie, la précision dépend du rapport $\frac{k}{h^2} \leqslant$ constante, si bien que pour un même pas de temps k et deux pas d'espace $h_1 \leqslant h_2$, le schéma correspondant à h_1 peut se trouver moins précis que celui qui correspond à h_2.

Tableau 28

S C H E M A I M P L I C I T E E N E Q U A T I O N												
		$\frac{1}{h}$ ECART$_1$										
h	k	t=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.	1.1
$\frac{1}{20}$	2.10^{-3}	.77	.98	.64	.23	.11	.08	.04	.05	.12	.37	.78
	10^{-3}	.61	.82	.67	.39	.21	.09	.04	.04	.11	.30	.61
$\frac{1}{40}$	2.10^{-3}	3.3	3.9	3.5	2.7	ARRET SUR TEMPS LIMITE						

Commentaires : Les données communes sont $f = -\frac{1}{2}(x - a(t))$, $x > a(t)$;

Test d'arrêt pour la méthode de Gauss-Seidel : 10^{-6} ; $f^n_{kh} = f^{n(\frac{1}{2})}_{kh}$.

f) Précision du schéma semi-implicite en projection

On donne dans le tableau suivant (Tableau 29) l'évolution de l'ECART entre la solution exacte et la solution approchée pour quatre valeurs du pas de temps : $k = 5.10^{-3}$, 2.10^{-3} , 10^{-3} , $0.5.10^{-3}$.

La meilleure valeur de k semble correspondre à $k = 10^{-3}$ ou $k = 0.5.10^{-3}$.

Tableau 29

S C H E M A S E M I - I M P L I C I T E E N P R O J E C T I O N											
	$\frac{1}{h}$ E C A R T$_1$										
k	t=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.	1.1
5.10^{-3}	0.50	0.24	1.33	1.50	0.72	0.20	0.06	0.05	0.10	0.25	0.51
2.10^{-3}	0.82	0.90	0.46	0.22	0.15	0.08	0.04	0.05	0.13	0.40	0.83
10^{-3}	0.28	0.54	0.82	0.75	0.38	0.13	0.04	0.04	0.08	0.16	0.28
$0.5.10^{-3}$	0.36	0.60	0.76	0.65	0.34	0.12	0.04	0.04	0.09	0.20	0.36

Commentaires : Les données communes sont $h = \frac{1}{20}$; $f^n_{h,k} = f^{n(+)}_{h,k}$;

$f = 0$ pour $x > a(t)$; Test d'arrêt pour la méthode de Gauss-Seidel : 10^{-6}.

g) Précision du schéma totalement implicite en projection

Dans le tableau suivant (Tableau 30) on donne l'évolution de l'ECART entre la solution exacte et la solution approchée pour trois valeurs du pas de temps.

Tableau 30

S C H E M A T O T A L E M E N T I M P L I C I T E											
	$\frac{1}{h}$ E C A R T_1										
k	t=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.	1.1
$2*10^{-3}$	0.41	0.73	0.86	0.57	0.26	0.10	0.04	0.04	0.09	0.20	0.41
10^{-3}	0.43	0.69	0.78	0.56	0.28	0.10	0.04	0.04	0.09	0.21	0.43
$0.5*10^{-3}$	0.44	0.67	0.74	0.56	0.28	0.11	0.04	0.04	0.10	0.22	0.44

Commentaires : Les données communes sont $h = \frac{1}{20}$; $f^n_{h,k} = f^{n(+)}_{h,k}$;

$f = -\frac{1}{2}(x - a(t))$ pour $x > a(t)$; Test d'arrêt pour la méthode de Gauss-Seidel: 10^{-6}

Des résultats - non conduits jusqu'à t = 1 - montrent par ailleurs que l'ECART s'accroit pour $k > 2*10^{-3}$ et $k < 0.5*10^{-3}$, la meilleure valeur semble donc se situer pour $k = 10^{-3}$.

h) Précision relative et rapidité des méthodes

Nous avons retenu pour cette comparaison les meilleurs résultats fournis par les divers schémas. Nous donnons le temps de résolution et l'évolution de l'$ECART_1$ entre les solutions approchées et la solution exacte (Tableau 31).

Tableau 31

COMPARAISON DES DIVERS SCHEMAS

			$\frac{1}{h}$ ECART$_1$											
Schémas	Caractéristiques		t=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.	1.1	Temps
S1	$h=\frac{1}{20}$	$k = 0.5.10^{-3}$	0.28	0.51	0.78	0.74	0.38	0.12	0.04	0.39	0.08	0.16	0.28	10mn529
S2		$k = 10^{-3}$	0.29	0.53	0.81	0.74	0.38	0.12	0.04	0.04	0.08	0.16	0.29	24mn
S3			0.27	0.54	0.82	0.74	0.38	0.13	0.04	0.04	0.07	0.16	0.28	24mn
S4			0.43	0.70	0.78	0.56	0.27	0.10	0.03	0.03	0.09	0.17	0.43	26mn

Commentaires : Les données communes sont $f = -\frac{1}{2}(x - a(t))$, pour $x > a(t)$;

$f^n_{h,k} = f^{n(+)}_{h,k}$; test d'arrêt pour la méthode de Gauss-Seidel : 10^{-6} .

S1 = Schéma explicite ,

S2 = Schéma implicite en équation ,

S3 = Schéma semi-implicite en projection ,

S4 = Schéma totalement implicite.

i) Conclusions

Pour la résolution de l'inéquation parabolique à contraintes réparties avec les données définies au début de ce paragraphe nous avons :

1. Choix de $f^n_{h,k}$
   - sans importance ,
   - nous conseillons cependant $f^n_{h,k} = f^{n(+)}_{h,k}$.
2. Schéma explicite
   - toujours le danger d'explosion ,
   - une évaluation correcte pour la stabilité semble être $\frac{k}{h^2} \leqslant \frac{1}{5}$,
   - à précision égale deux fois plus rapide que les autres schémas.
3. Les autres schémas
   - les schémas implicites en équation, semi-implicite en projection et totalement implicites semblent être ici équivalents. ∎

2.9. Un problème hyperbolique à contraintes réparties d'ordre 0

On considère, pour $t \in [0,T]$, l'inéquation

$$(u'' - \Delta u - f,\ v - u) \geqslant 0 \ , \ \forall v \in K,\ u \in \mathcal{K}^{(1)}$$

avec

$$V = H^1_o(\Omega) ,$$

$$K = \{ v \mid v \in V, \quad v \geqslant 0, \text{ p.p. sur } \Omega \} ,$$

$$\mathcal{K} = \{ v \mid v \in L^2(0,T;V),\ v' \in L^2(0,T;V),\ v'' \in L^2(0,T;V'),$$
$$v(t,x) \geqslant 0 \text{ , p.p. sur } Q =]0,T[\times \Omega ,$$
$$v(o,x) = u_o \text{ donné, } v'(o,x) = u_1 \text{ donné.} \} .$$

En prenant

$$\Omega =]0,1[\times]0,1[,$$

$$a(t) = \frac{1}{4}(1 + 3t) ,$$

$$f = \begin{cases} -\frac{1350}{4} x(x - a(t))\, y(1 - y) - 200x(x - a(t))^3 + \\ + 600(x - a(t))\,(2x - a(t))\, y(1 - y) \quad \text{si } x \leqslant a(t) ; \\ 0 \text{ quelconque si } x > a(t) , \end{cases}$$

(1) $u \in \mathcal{K}$ et non $u' \in \mathcal{K}$.

le problème possède la solution exacte

$$u = \begin{cases} -100\, x\, (x - a(t))^3\, y(1-y) & \text{si } x \leq a(t), \\ 0 & \text{si } x > a(t), \end{cases}$$

avec pour conditions initiales

$$\begin{cases} u_o = u(t=0), \\ u_1 = u'(t=0). \end{cases}$$

. On a représenté u en fonction du temps sur le graphique ci-après (Figure 35). On donne sur la Figure 36 quelques équipotentielles de la fonction u aux instants t = 0.2 et t = 0.9.

a) Description des schémas

La discrétisation en espace est la discrétisation classique de $H_o^1(\Omega)$.

Pour la discrétisation en temps nous avons retenu les schémas suivants :

Schéma explicite

$$\begin{cases} u_h^{-1}\ ,\ \delta_h^o \text{ donnés et } u_h^o = u_h^{-1} + k\delta_h^o\ , \\ \dfrac{\delta_h^{n+1} - \delta_h^n}{k} - \Delta_h u_h^n - f_h^n = 0\ , \\ u_h^{n+1} = \max\left[u_h^n + k\,\delta^{n+1}\ ,\ 0 \right]. \end{cases}$$

Schéma implicite en équation

$$\begin{cases} u_h^{-1},\ \delta_h^o \text{ donnés et } u_h^o = u_h^{-1} + k\delta_h^o. \text{ On pose } \delta_h^n = \dfrac{u_h^n - u_h^{n-}}{k} \\ \text{puis} \\ \dfrac{\delta_h^{n+\frac{1}{2}} - \delta_h^n}{k} - \Delta_h u_h^{n+\frac{1}{2}} - f_h^n = 0 \text{ avec } \delta_h^{n+\frac{1}{2}} = \dfrac{u_h^{n+\frac{1}{2}} - u}{k} \\ u_h^{n+1} = \max\left[u_h^{n+\frac{1}{2}}\ ,\ 0 \right]. \end{cases}$$

On peut encore écrire ce schéma comme suit

$$\begin{cases} \dfrac{\delta_h^{n+\frac{1}{2}} - \delta_h^n}{k} - \Delta_h(u_h^n + k\,\delta_h^{n+\frac{1}{2}}) - f_h^n = 0\ , \\ u_h^{n+1} = \max\left[u_h^n + k\,\delta_h^{n+\frac{1}{2}}\ ,\ 0 \right]\ , \end{cases}$$

Figure 35

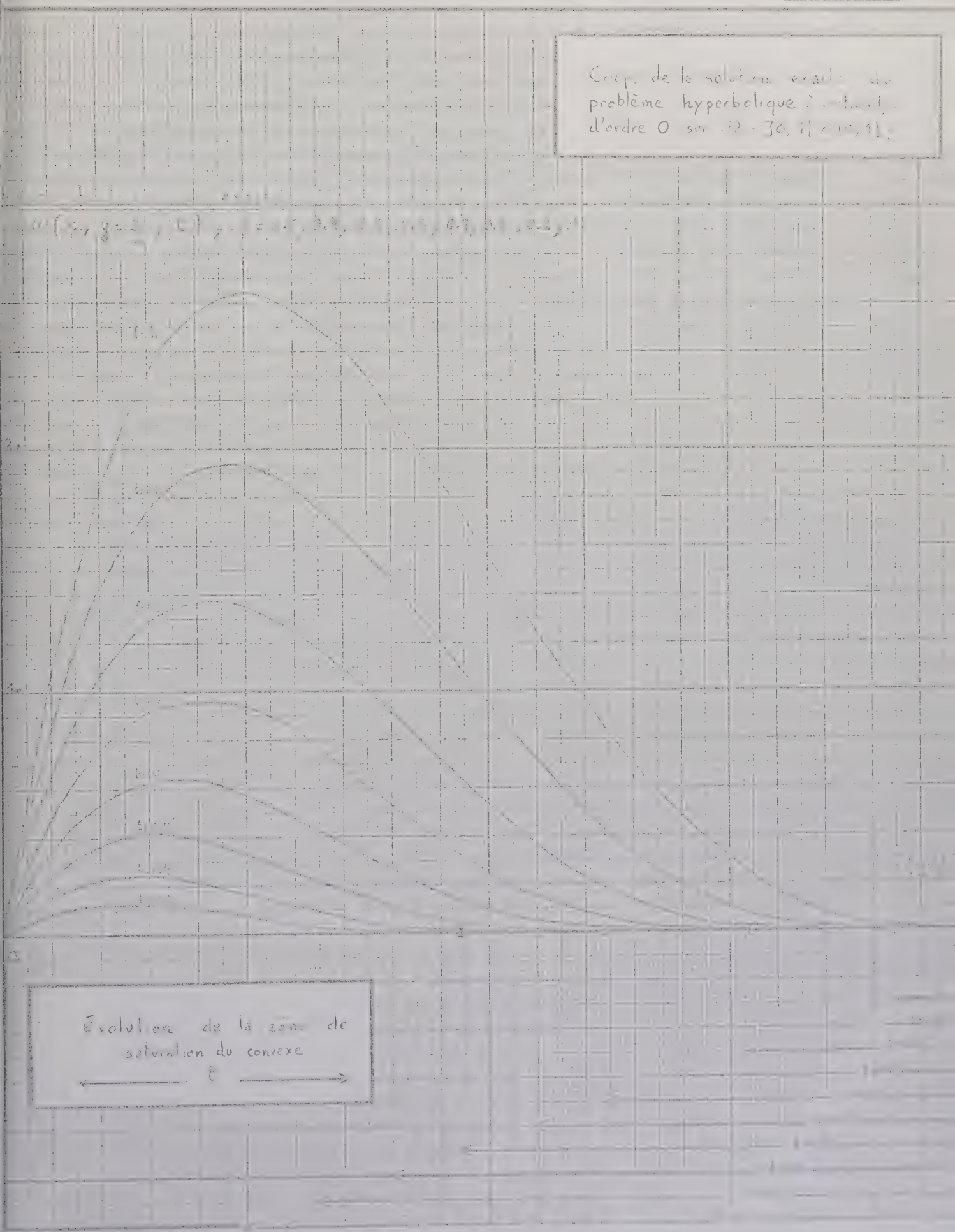


Figure 36

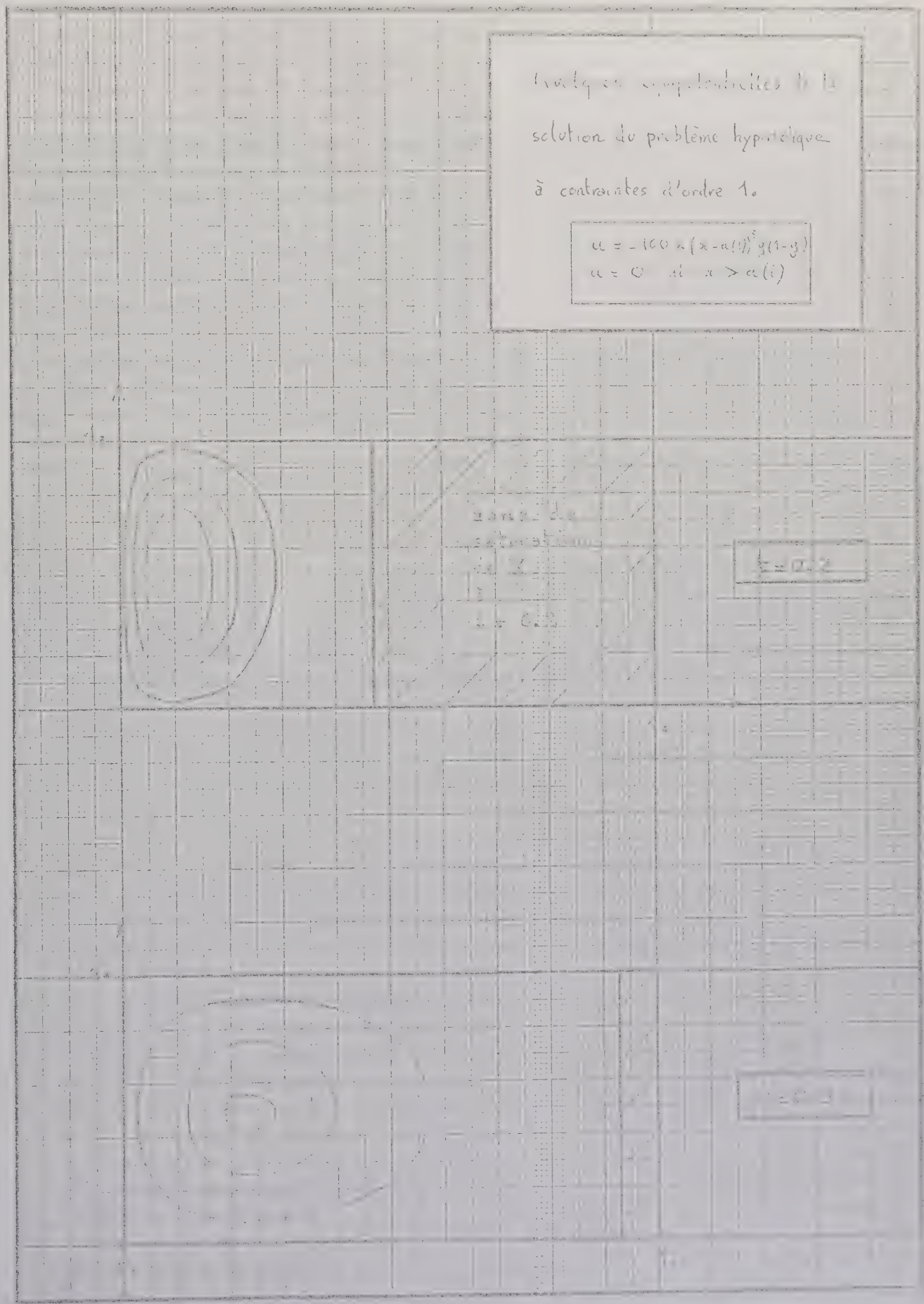

Quelques compatibilités de la solution du problème hyperbolique à contraintes d'ordre 1.
$u = -100\,x\,(x-a(i))^5\,y(1-y)$
$u = 0 \quad si \quad x > a(i)$
zone de calcul
$t = 0.2$
$t = 0.3$

Figure 37

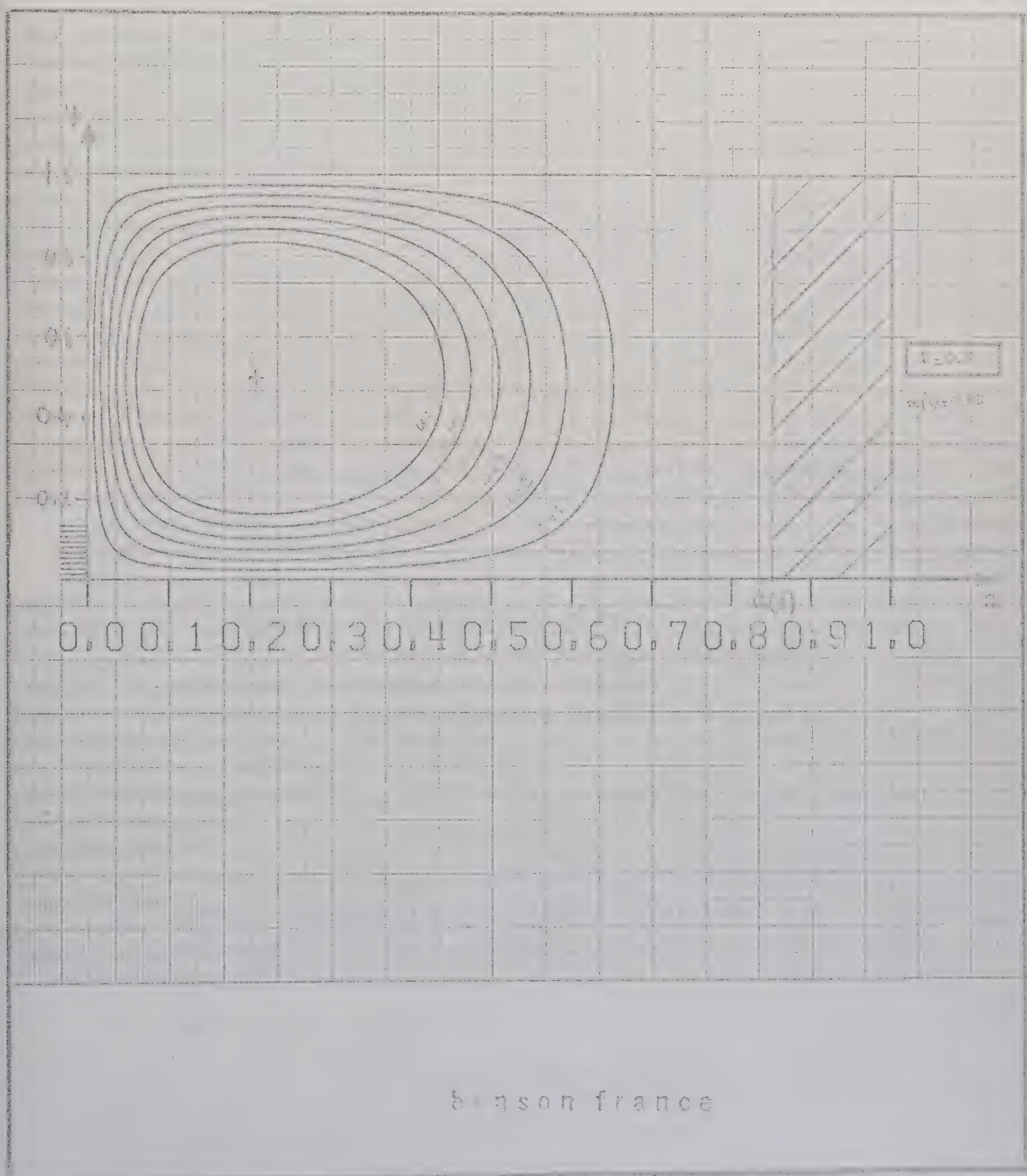


Schéma semi-implicite en projection

$$\left|\begin{array}{l} u_h^{-1}\ ,\ \delta_h^o \text{ donnés et } u_h^o = u_h^{-1} + k\,\delta_h^o\ , \\ \\ \left(\dfrac{\delta_h^{n+1} - \delta_h^n}{k} - \Delta_h\left(\dfrac{u_h^n + u_h^{n+1}}{2}\right) - f_h^n\ ,\ v - u_h^{n+1}\right)_h \geqslant 0, \\ \\ \forall\, v \geqslant 0,\ u_h^{n+1} \geqslant 0\ . \end{array}\right.$$

Ce que l'on peut encore écrire

$$\left|\begin{array}{l} \left(\dfrac{u_h^{n+1} - 2u_h^n + u_h^{n-1}}{k^2} - \Delta_h\left(\dfrac{u_h^n + u_h^{n+1}}{2}\right) - f_h^n\ ,\ v - u_h^{n+1}\right)_h \geqslant 0 \\ \\ \forall\, v \geqslant 0,\ u_h^{n+1} \geqslant 0. \end{array}\right.$$

Schéma totalement implicite

$$\left|\begin{array}{l} u_h^{-1}\ ,\ \delta_h^o \text{ donnés et } u_h^o = u_h^{-1} + k\,\delta_h^o\ , \\ \\ \left(\dfrac{\delta_h^{n+1} - \delta_h^n}{k} - \Delta_h\, u_h^{n+1} - f_h^n\ ,\ v - u_h^{n+1}\right)_h \geqslant 0, \\ \\ \forall\, v \geqslant 0,\ u_h^{n+1} \geqslant 0. \end{array}\right.$$

b) Résultats des différents schémas

On donne dans le tableau suivant (Tableau 32) les résultats pour les schémas décrits en a).

On a pris pour conditions initiales discrètes

$$u_h^{-1} = 0\ ,\quad \delta_h^o(M) = u'(o,M)\ ,$$

ce qui est, pour u_h^{-1}, une approximation à 5.10^{-3} des valeurs réelles de la condition initiale.

Pour les schémas autres que le schéma explicite, la méthode utilisée à chaque itération en temps est la méthode de Gauss-Seidel ou la méthode de Gauss-Seidel en projection avec une approximation du paramètre optimal $\omega_{opt.théorique}$ donnée par la formule :

$$\omega = \frac{2}{1 + \sqrt{1 - \dfrac{2\left(1 - \dfrac{\lambda_1^2}{2}h^2\right)h^2}{2 + \dfrac{h^2}{k^2}}}}\ .$$

Le test d'arrêt de la méthode de Gauss-Seidel a été fixé à

$$\sum_{M \in \Omega_h} \left| u_h^{n+1(i+1)}(M) - u_h^{n+1(i)} \right| < 10^{-5}$$

où l'on note $u_h^{n+1(i)}$ la solution trouvée à l'itération i de la méthode. Le pas d'espace a été fixé à $h = \frac{1}{20}$.

Pour $x > a(t)$ on a pris $f = -\frac{1}{2}(x - a(t))$.
La programmation a été faite en Fortran H sur CII-10070.

Tableau 32

INEQUATION HYPERBOLIQUE A CONTRAINTES REPARTIES D'ORDRE 0												
$h = \frac{1}{20}$			$\frac{1}{h}$ ECART$_1$ (solution exacte - solution trouvée)									
Schémas	k	Temps	t=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1
Explicite	5.10^{-4}	/	16.	EXPLOSION									
	10^{-4}	38mn	.48	.43	.38	.35	.37	.39	.43	.47	.53	.61	/
Implicite en équation	10^{-3}	13mn	.49	.43	.37	.33	.30	.33	.41	.53	.74	1.02	1.16
Semi-Implicite en équation	10^{-2}	3mn	.48	.40	.47	.74	1.20	1.86	2.81	4.06	5.63	7.47	/
	5.10^{-3}	4mn	.48	.42	.41	.48	.66	.94	1.29	1.77	2.43	3.20	/
	2.10^{-3}	7mn	.48	.43	.39	.36	.37	.40	.45	.49	.56	.67	/
	10^{-3}	13mn	.48	.43	.38	.32	.30	.29	.30	.34	.43	.57	/
Totalement Implicite	10^{-3}	14mn	.49	.43	.37	.33	.30	.33	.41	.53	.73	1.02	1.17

Commentaires : Les barres obliques indiquent que les valeurs correspondantes n'ont malheureusement pas été demandées.

c) Conclusions

Pour le problème hyperbolique à contraintes réparties d'ordre 0 traité ici nous avons :

1. Schéma explicite
   - toujours le danger d'explosion,
   - une évaluation correcte pour la stabilité est $\frac{k}{h} \leqslant 2.10^{-3}$.

2. Les autres schémas
   - sensiblement équivalents,
   - à précision égale, tous de 2 à 3 fois plus rapides que le schéma explicite,
   - une valeur admissible de $\frac{k}{h}$ est 2.10^{-2}.

2.10. Un problème hyperbolique à contraintes réparties d'ordre 1

On considère pour $t \in [0,T]$, l'inéquation

$$(u'' - \Delta u - f,\ v - u') \geqslant 0, \quad \forall v \in K, \quad u \in \mathcal{K},$$

avec

$$V = H_0^1(\Omega),$$

$$K = \{ v \mid v \in V, \quad v \geqslant 0, \text{ p.p. sur } \Omega \},$$

$$\mathcal{K} = \{ v \mid v \in L^2(0,T;V),\ v' \in L^2(0,T;V),\ v'' \in L^2(0,T;V'),$$
$$v(o) = u_o \text{ donné},\ v'(o) = u_1 \text{ donné},$$
$$v' \geqslant 0, \text{ p.p. sur } Q =]0,T[\times \Omega \}.$$

Nous prenons $\Omega =]0,1[\times]0,1[$ et nous définissons f, u_o, u_1 de façon à avoir pour solution exacte

$$u = \begin{cases} 10 \sin\left(\frac{3\pi}{2} t\right) x(1-x)\, y(1-y), & 0 \leqslant t \leqslant \frac{1}{3}, \\ 10\, x(1-x)\, y(1-y), & \frac{1}{3} < t < \frac{2}{3}, \\ 10\left(1 - \sin\frac{3\pi}{2} t\right) x(1-x)\, y(1-y), & \frac{2}{3} \leqslant t \leqslant 1 = T. \end{cases}$$

La solution (voir Figure 37) "monte" de $t = 0$ à $\frac{1}{3}$, se stabilise entre $\frac{1}{3}$ et $\frac{2}{3}$ et "remonte" pour $t = \frac{2}{3}$ jusqu'à 1.

Afin que ce problème se présente réellement sous la forme d'une inéquation (et non d'une équation variationnelle),

nous prenons

$$f = \begin{cases} u'' - \Delta u \ , & 0 \leqslant t \leqslant \frac{1}{3} \ , \\ - \Delta u - (t - \frac{1}{3})(\frac{2}{3} - t) \ , & \frac{1}{3} < t < \frac{2}{3} \ , \\ u'' - \Delta u \ , & \frac{2}{3} \leqslant t \leqslant 1 \ . \end{cases}$$

a) Description des schémas

La discrétisation en espace est la discrétisation classique de $H_o^1(\Omega)$.

Pour la discrétisation en temps nous avons retenu les schémas suivants :

Schéma explicite

$$\left| \begin{array}{l} u_h^{-1} \ , \ \delta_h^o \text{ donnés et } u_h^o = u_h^{-1} + k \delta_h^o \ , \\ \dfrac{\delta_h^{n+\frac{1}{2}} - \delta_h^n}{k} - \Delta_h u_h^n - f_h^n = 0 \ , \\ \delta^{n+1} = \max\left[\delta^{n+\frac{1}{2}} , 0 \right] \ , \\ u^{n+1} = u^n + k \delta^{n+1} \ . \end{array} \right.$$

Schéma implicite en équation

$$\left| \begin{array}{l} u_h^{-1} \ , \ \delta_h^o \text{ donnés et } u_h^o = u_h^{-1} + k \delta_h^o \ , \\ \dfrac{\delta_h^{n+\frac{1}{2}} - \delta_h^n}{k} - \Delta_h u_h^{n+\frac{1}{2}} - f_h^n = 0 \text{ avec } u_h^{n+\frac{1}{2}} = u_h^n + k\delta_h^{n+} \\ \delta^{n+1} = \max\left[\delta^{n+\frac{1}{2}} , 0 \right] \ , \\ u^{n+1} = u^n + k \delta^{n+1} \ . \end{array} \right.$$

Schéma semi-implicite en projection

$$\left| \begin{array}{l} u_h^{-1} \ , \ \delta_h^o \text{ donnés et } u_h^o = u_h^{-1} + k \delta_h^o \ , \\ \left(\dfrac{\delta_h^{n+1} - \delta_h^n}{k} - \Delta_h \left(\dfrac{u_h^n + u_h^{n+1}}{2} \right) - f_h^n , v - \delta_h^{n+1} \right) \geqslant 0, \\ \forall v \geqslant 0, \quad \delta_h^{n+1} \geqslant 0 \end{array} \right.$$

avec $u_h^{n+1} = u_h^n + k \delta_h^{n+1}$.

Figure 37

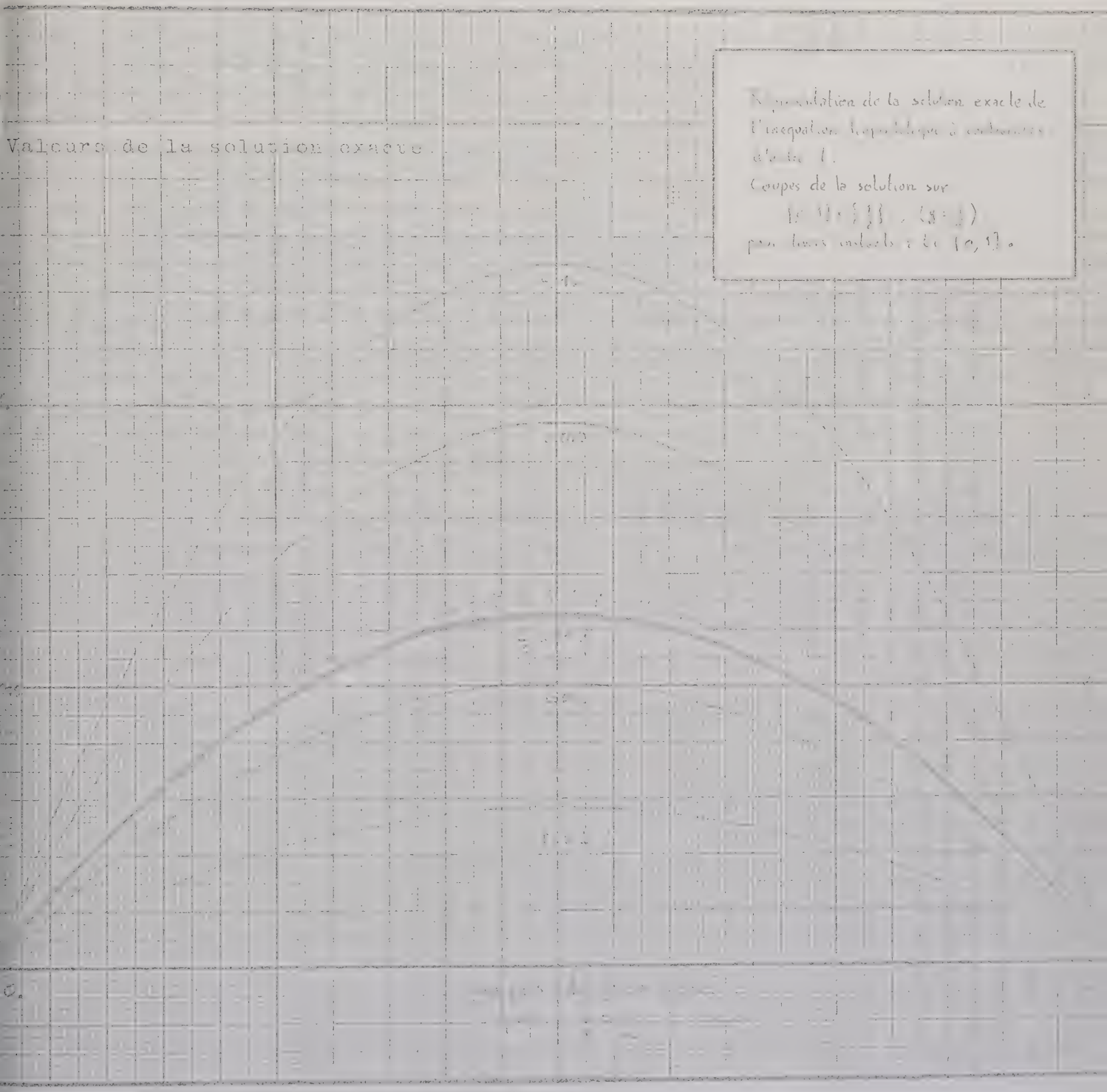


Schéma implicite en projection

$$
\left|
\begin{array}{l}
u_h^{-1},\ \delta_h^o \text{ donnés et } u_h^o = u_h^{-1} + k\,\delta_h^o , \\
\left(\dfrac{\delta_h^{n+1} - \delta_h^n}{k} - \Delta_h u_h^{n+1} - f_h^n ,\ v - \delta_h^{n+1} \right) \geq 0, \\
\forall\, v \geq 0, \quad \delta_h^{n+1} \geq 0 \\
\text{avec } u_h^{n+1} = u_h^n + k\,\delta_h^{n+1} .
\end{array}
\right.
$$

b) Résultats des différents schémas

On donne dans le tableau suivant (Tableau 33) les résultats pour les schémas décrits en a).

On a pris pour conditions initiales discrètes

$$
\left|
\begin{array}{l}
u_h^o = u(o) = 0 , \\
\delta_h^o(M) = u'(o,M) .
\end{array}
\right.
$$

Pour les schémas autres que le schéma explicite, la méthode utilisée à chaque pas de temps est la méthode de Gauss-Seidel ou la méthode de Gauss-Seidel en projection avec la même approximation du paramètre ω que celle qui est donnée au § 2.9 précédent. Le test d'arrêt est

$$
\sum_{M \in \Omega_h} \left| \delta_h^{n+1(i+1)}(M) - \delta_h^{n+1(i)} \right| < 10^{-5}
$$

où $\delta_h^{n+1(i)}$ est la solution trouvée à l'itération i de la méthode. Le pas d'espace a été fixé à $h = \frac{1}{20}$.

La programmation a été faite en Fortran H sur CII 10070.

Tableau 33

INEQUATION HYPERBOLIQUE A CONTRAINTES REPARTIES D'ORDRE 1												
$h = \frac{1}{20}$			$\frac{1}{h}$ ECART$_1$ (solution exacte - solution trouvée)									
Schémas	k	Temps	t=0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.	1.1
Explicite	10^{-3}	4mn	2.05	7.47	14.14	16.68	16.68	16.68	16.59	14.77	12.19	10.6	/
	5.10^{-4}	16mn	1.00	3.63	7.12	9.71	9.71	9.71	9.62	8.02	6.75	4.58	/
	2.10^{-4}	30mn	0.39	1.42	2.63	3.00	3.00	3.00	2.89	2.73	2.66	2.12	/
	10^{-4}	57mn	0.19	0.70	1.19	1.48	1.48	1.48	1.43	1.38	1.31	1.02	/
Implicite en équation	10^{-3}	16mn	0.05	0.19	0.36	0.41	0.41	0.41	0.31	0.01	0.23	0.32	0.24
Semi-implicite en inéquation	10^{-3}	16mn	0.02	0.10	0.19	0.22	0.22	0.22	0.12	0.16	0.37	0.46	0.40
Implicite en inéquation	10^{-3}	16mn	0.05	0.19	0.36	0.41	0.41	0.41	0.31	0.01	0.23	0.32	0.24

c) Conclusions

Pour le problème hyperbolique à contraintes réparties d'ordre 1 que nous venons de traiter nous avons:

1. Schéma explicite
   - toujours le danger d'explosion ,
   - la stabilité semble correspondre à $\frac{k}{h} \leqslant 2.10^{-3}$.

2. Les autres schémas
   - sensiblement équivalents ,
   - une valeur admissible de $\frac{k}{h}$ est 2.10^{-2} ,
   - à précision égale ces schémas sont 3 fois plus rapides que le schéma explicite.

2.11. Conclusion générale

Les exemples que nous avons traités ne peuvent donner qu'une faible indication sur ce qu'il convient de faire lorsqu'il faut résoudre un problème réel. Ils nous permettent cependant de nous faire une idée sur les méthodes et schémas à utiliser face aux situations concrètes. Pour préciser cela nous allons reprendre très rapidement les divers problèmes abordés dans ce dernier chapitre.

Problèmes Stationnaires Quadratiques, Différentiables et sans Contraintes

De toutes les méthodes traitées, la méthode la plus intéressante est celle du gradient conjugué, en temps, en nombre d'itérations et en précision.

Cette méthode est cependant plus délicate à programmer que la méthode de Gauss-Seidel avec paramètre de relaxation optimal. Celle-ci est très aisée à mettre en oeuvre, mais la détermination du paramètre de relaxation peut être délicate (et ce paramètre est à calculer pour chaque problème : il est obtenu sans difficulté pour le problème de Neumann, mais beaucoup plus difficile à calculer avec précision pour le problème de Dirichlet, fort heureusement une sous-évaluation approximative, dans le cas de ce dernier problème, est suffisante ; pour le problème de Neumann, le paramètre doit être calculé avec une très bonne précision).

Problèmes Stationnaires non Différentiables

Les méthodes expérimentées sont sensiblement équivalentes.

Problèmes Stationnaires avec Contraintes de Bornes

Les méthodes S.N. et de Gauss-Seidel sont sensiblement équivalentes, cette dernière étant d'autant plus intéressante qu'il y a plus de contraintes ; elle est de plus excessivement simple à programmer.

Problèmes Stationnaires avec Contraintes non Linéaires, ou "Obliques"

La seule méthode présentée est celle des centres à troncature variable. Cette méthode est lente et assez délicate à

programmer, de même d'ailleurs que toutes les autres méthodes de pénalisation intérieure. On préfèrera utiliser, lorsque le problème s'y prête, (et par difficulté croissante de programmation), une méthode de dualité, de pénalisation externe, d'approches polyédriques, ou de gradient projeté.

Problèmes d'Evolution Parabolique et Hyperbolique

Si l'objectif est de déterminer la solution en un très grand nombre d'instants très rapprochés, on préfèrera un schéma explicite. Mais ce genre de schéma est dangereux et pas toujours plus rapide que les schémas implicites. Ces derniers sont extrêmement stables, sont très précis et dans tous les cas préférable aux multiples variantes, schémas semi-implicites ou implicites en équation, semi-implicites en inéquation. Nous pensons que les schémas totalement implicites sont essentiellement valables lorsqu'on désire avoir connaissance de la solution à des instants relativement espacés. Bien que cela n'ait pas été fait ici, les schémas implicites sont dans tous les cas préférables[1] aux schémas explicites lorsque l'on peut, dès le départ, inverser la matrice du système une fois pour toute (il faut cependant que cela soit effectué avec précision).

Pour définir les schémas de façon correcte nous pensons qu'il est mieux de passer d'abord à la formulation de l'énergie, de discrétiser le potentiel, puis de définir les conditions d'Euler définissant le minimum de l'énergie. Cela évite certaines erreurs qui sont fatales dans ce genre de problème et malheureusement pas toujours faciles à détecter.

En l'absence de connaissance de solution exacte (!), le seul moyen de contrôler la valeur d'un schéma est de le résoudre pour deux couples (k_o , h_o) et (k_1 , h_1) avec $k_1 < k_o$, $h_1 < h_o$ et

$$\frac{k_1}{h_1^2} < \frac{k_o}{h_o^2} \quad \text{(pour les problèmes paraboliques)},$$

$$\frac{k_1}{h_1} < \frac{k_o}{h_o} \quad \text{(pour les problèmes hyperboliques)}.$$

(1) Cela n'ayant de sens que dans la résolution des équations.

et de regarder si les solutions obtenues sont suffisamment proches l'une de l'autre.

Dans toutes les méthodes nous conseillons d'utiliser comme critère d'arrêt la norme L_1 ou L_2 du gradient de la fonction d'énergie.

B I B L I O G R A P H I E

K.J. ARROW, L. HURWICZ et H. UZAWA[1],
Studies in Linear and Non Linear Programming, Stanford (1958).

J.P. AUBIN[1],
Approximation des Espaces de Distributions et des Opérateurs Différentiels. Thèse, Paris, (1966).

C. BARDOS et H. BREZIS[1],
Sur une Classe de Problèmes d'Evolution Non Linéaires. J. of. Diff. équations, Vol. 6, 345-394, (1969).

F.S. BECKMAN[1],
The Solution of Linear Equations by the Conjugate Gradient Method. Dans Mathematical Methods, Ralston et Wilf ed., Wiley, 62-72, (1960).

C. BERGE[1],
Espaces Topologiques, Fonctions Multivoques. Dunod, Paris, (1959).

N. BOURBAKI[1],
Intégration. Hermann, Paris, (1952).

J.F. BOURGAT[1],
Analyse Numérique du Problème de la Torsion Elastoplastique. Thèse 3e cycle, Paris, (1971).

J.F. BOURGAT - R. TREMOLIERES [1],
Refroidissement d'un Cylindre Radioactif (à paraître, Note IRIA).

H. BREZIS[1],
Equations et Inéquations Non Linéaires dans les Espaces Vectoriels en Dualité. Ann. Inst. Fourier, Grenoble, Vol. 18, No. 1, 115-175, (1968).

H. BREZIS[2],
Problèmes Unilatéraux. Thèse, Paris, (1971).

H. BREZIS et M. SIBONY [1],
Equivalence de Deux Inéquations Variationnelles et Applications. (Déc. 1969).

H. BREZIS et M. SIBONY [1],
Méthodes d'Approximation et d'Itération pour les Opérateurs Monotones.

H. BREZIS et STAMPACCHIA [1],
Sur la Régularité de la Solution d'Inéquations Elliptiques. Bull. Soc. Math. France, 96, 153-180, (1968).

F.E. BROWDER[1],
Assymptotic Fixed Point Theorems. Math. Ann. 185, 38-60, (1970).

F.E. BROWDER[2],
Existence and Approximation of Solutions of Non Linear Variational Inequalities. Proc. N.A.S., Vol. 56, 1080-1086, (1966).

F.E. BROWDER[3],
Fixed-Point Theorems for Non Compact Mappings in Hilbert Space. Proc. N.A.S., Vol. 53, 1272-1276, (1965).

F.E. BROWDER[4],
Mapping Theorems for Non Compact Non Linear Operators in Banach Space. Proc. N.A.S., Vol. 54, No. 2, 337-342, (Aug. 15, 1965).

F.E. BROWDER[5],
Non Linear Operators in Banach Spaces. Math. Annalen 162, 280-283, (1966).

F.E. BROWDER[6],
Non Linear Maximal Monotone Operators in Banach Space. Math. Annalen 175, 89-113, (1968).

F.E. BROWDER[7],
Non Linear Variational Inequalities and Maximal Monotone Mappings in Banach Spaces. Math. Ann. 183, 213-231, (1969).

F.E. BROWDER[8],
On the Unification of the Calculus of Variations and the Theory of Monotone Non Linear Operators in Banach Spaces. Proc. N.A.S., Vol. 56, 419-425, (1966).

F.E. BROWDER[9],
The Fixed Point Theory of Multi-Valued Mappings in Topological Vector Spaces. Math. Annalen 177, 283-301, (1968).

F.E. BROWDER et W.V. PETRYSHIN[1],
Construction of Fixed Points of Non Linear Mappings in Hilbert Space. J. of. Math. Anal. and Appl., Vol. 20, 197-228, (1967).

J. CEA[1],
Approximation Variationnelle des Problèmes aux Limites. Thèse, Paris, (1965).

J. CEA[2],
Optimisation : Théorie et Algorithmes. Dunod, Paris, (1971).

J. CEA et GOULAOUIC[1],
C_4 Analyse Fonctionnelle et Numérique. Renne, (1967-1968).

G. DUVAUT et J.L. LIONS[1],
Les Inéquations en Mécanique et en Physique. Dunod, Paris, (1972).

A.M. GEOFFRION[1],
Duality in Non Linear Programming: A Simplified Applications-Oriented Development. SIAM Review, Vol. 13, No. 1, 1-37, (Jan. 1971).

A. GHIZZETTI, ed. [1],
Theory and Applications of Monotone Operators. ODERISI, Gubbio, (1969).

GLOWINSKI - LIONS - TREMOLIERES [1],
Méthodes Numériques de Résolution des Problèmes d'Inéquations en Mécanique et en Physique. Dunod, Paris, (1972) (à paraître).

M. GOURSAT [1],
Analyse Numérique de Problèmes d'Elastoplasticité et de Viscoelasticité. Thèse 3e cycle, Paris, (1971).

P. HUARD [1],
Method of Centers. Ch. 7 in Non Linear Programming. ed. J. Abadie, John Wiley & Sons, Inc., New York, (1967).

KOLMOGOROV et FOMIN [1],
Functional Analysis. Graylock, (1961).

KY-FAN [1],
Fixed Point and Minimax Theorems in Locally Convex Topological Linear Spaces. Proc. N.A.S., Vol. 38, (1952).

J.L. LIONS [1],
Equations Différentielles Opérationnelles. Springer-Verlag, Berlin, (1961).

J.L. LIONS [2],
Méthodes d'Approximation Numérique des Problèmes aux Limites de la Physique Mathématique. C.N.R.S., (Cours 1961-1962).

J.L. LIONS [3],
Optimisation pour Certaines Classes d'Equations d'Evolution Non Linéaires. Annali di Mat. Pura ed Applicata (IV), Vol. 72, 275-294, (1966).

J.L. LIONS [4],
Problèmes aux Limites dans les Equations aux Dérivées Partielles. Cours de Montréal, (1962).

J.L. LIONS [5],
Quelques Méthodes de Résolution des Problèmes aux Limites Non Linéaires. Dunod, Paris, (1969).

J.L. LIONS [6],
Sur de Nouveaux Problèmes d'Inéquations Variationnelles Posés par la Mécanique. Le Cas d'Evolution. CRAS, t. 269, 570-572, (6 oct. 1969).

J.L. LIONS [7],
Sur l'Approximation de la Solution d'Inéquations d'Evolution. CRAS, t. 203, 55-57, (11 juillet 1966).

J.L. LIONS - E. MAGENES [1],
Problèmes aux Limites non Homogènes et Applications. Dunod, Paris, (1968).

J.L. LIONS et P.A. RAVIART [1],
Remarques sur la Résolution, Exacte et Approchée, d'Equations d'Evolution Paraboliques à Coefficients non Bornés. Calcolo, Vol. 4, fax. 2, 221-234, (1967).

J.L. LIONS et G. STAMPACCHIA [1],
Variational Inequalities. Communications on Pure and Applied Mathematics. Vol. XX, 493-519, (1967).

D.G. LUENBERGER[1],
Optimization by Vector Space Methods. Wiley, (1969).

O.L. MANGASARIAN et J. PONSTEIN [1],
Minimax and Duality in Non Linear Programming. Journal of Mathematical Analysis and Applications, Vol. 11, No. 1-3, 504-18, (1965).

Ph. MORICE - R. TREMOLIERES [1],
Parallelisation et Eléments Finis (à paraître).

J.F. NASH[1],
Equilibrium Points in N-Person Games. Proc. N.A.S., Vol.36, 48-49, (1950).

J. NECAS[1],
Les Méthodes Directes en Théorie des Equations Elliptiques. Masson, Paris, (1967).

NEDELEC[1],
Cours de D.E.A. (Approximation), Rennes, (1968-1969).

P.A. RAVIART[1],
Approximation des Equations d'Evolution par des Méthodes Variationnelles. Rennes, (1966-1967).

P.A. RAVIART[2],
Cours d'Analyse Numérique. Rennes, (1966-1967).

P.A. RAVIART[3],
Sur l'Approximation de Certaines Equations d'Evolution Linéaires et Non Linéaires. Thèse, Paris, (1965).

RIESZ et NAGY [1],
Leçons d'Analyse Fonctionnelle. Paris, Gauthier-Villars, (1968).

J.B. ROSEN[1],
The Gradient Projection Method for Non Linear Programming. Part I. Linear Constraints. J. SIAM, Vol. 8, No. 1, 181-217, (March 1969).

J.B. ROSEN [2],
The Gradient Projection Method for Non Linear Programming. Part II. Non Linear Constraints. J. SIAM, Vol. 9, No. 4, 514-532, (Dec. 1961).

R. TENAM[1],
Approximation de Problèmes aux Limites Elliptiques. C_3 Analyse Numérique. Orsay, (1970).

R. TENAM [2],
Sur la Stabilité et la Convergence de la Méthode des Pas Fractionnaires. Thèse, Paris, (1968).
Annali di Matematica Pura ed Applicata (IV), Vol. 79, 191-380, (1968).

F. THOMASSET [1],
Etude Comparative de quelques Méthodes de Minimisation.
I.R.I.A., non publié, (1971).

R. TREMOLIERES [1],
Algorithmes pour la Résolution des Problèmes Posés Sous La Forme Min Max f(x,y). Note I.R.I.A., Inf. 6909, (15 mai 1969).

R. TREMOLIERES [2],
Analyse Fonctionnelle et Approximation des Inéquations. Cours C_3 Toulouse, (1970-1971).

R. TREMOLIERES [3],
Approximation Externe du Problème de la Torsion Elasto-Plastique d'une Barre. Note I.R.I.A., Inf/7016, (fév. 1970).

R. TREMOLIERES [4],
Inéquations Non Symétriques. Note I.R.I.A. (1971).

R. TREMOLIERES [5],
La Méthode des Centres à Troncature Variable. Thèse 3e cycle, Paris, (1968).

R. TREMOLIERES [6],
Méthode des Eléments Finis et de Ritz-Galerkin pour les Problèmes à Frontière Libre. (Non publié).(janvier 1971).

R. TREMOLIERES [7],
Optimisation Non Linéaire. (à paraître).

R. TREMOLIERES [8],
Théorie et Méthodes de la Dualité. Note I.R.I.A., INF/7212/72015, (1972).

R. TREMOLIERES [9],
Principes Variationnels en Thermo-Elasto-Plasticité. Note I.R.I.A., Inf/7002, (fév. 1970).

R. TREMOLIERES - Ph. MORICE
Voir MORICE - TREMOLIERES.

R.S. VARGA [1],
Matrix Iterative Analysis. Prentice Hall, (1962).

D. VIAUD [1].
Thèse, (à paraître), Paris, (1972).

K. YOSIDA [1],
Functional Analysis. Springer-Verlag, (1966).

A.C. ZAANEN [1],
Linear Analysis. North-Holland, (1960).

E R R A T A :

Lire page 118 :

$$\omega_p^{n+1} = \frac{2}{1+\sqrt{1-\rho_g^{n+1}}} \quad , \qquad \omega_g^{n+1} = \frac{2}{1+\sqrt{1-\rho_g^{n+1}}}$$

on peut montrer que .

$$\rho_p^1 \leqslant \rho_p^2 \leqslant \ldots \leqslant \rho_p^n \leqslant \ldots \rho(\mathcal{L}_1) \ldots \leqslant \rho_g^n \leqslant \ldots \leqslant \rho_g^2 \leqslant \rho_g^1 \quad ,$$

$$\omega_p^1 \leqslant \omega_p^2 \leqslant \ldots \leqslant \omega_p^n \leqslant \ldots \omega_{opt} \ldots \leqslant \omega_g^n \leqslant \ldots \leqslant \omega_g^2 \leqslant \omega_g^1 \quad ,$$

et que

$$\rho_p^n - \rho_g^n \longrightarrow 0 \, , \qquad \omega_p^n - \omega_g^n \longrightarrow 0 \, ,$$

$$\rho_p^n \longrightarrow \rho(\mathcal{L}_1) \quad , \qquad \rho_g^n \longrightarrow \rho(\mathcal{L}_1) \quad ,$$

$$\omega_p^n \longrightarrow \omega_{opt} \quad , \qquad \omega_g^n \longrightarrow \omega_{opt} \quad .$$

et en particulier

$$\omega_{opt}^n = \frac{\omega_p^n + \omega_g^n}{2} \longrightarrow \omega_{opt} \quad .$$

Lire page 115 :

P_n au lieu de P_h

ACADÉMIE DE PARIS

UNIVERSITÉ DE PARIS VI

4, Place Jussieu - 75230 PARIS-CEDEX 05

SERVICE DE LA SCOLARITE

Bureau des Enseignements du 3e cycle

TÉLÉPHONE : 336-25-25 - 325-12-21

PERMIS D'IMPRIMER

DOCTORAT : . . DOCTORAT ES SCIENCES . . . . . .

THESE DE : . . . . MATHEMATIQUES . . . . . .

NOM : Monsieur TREMOLIERES . . . . . .

Prénoms : . . Raymond . . . . . .

né le 20 septembre 1941 à . . . . LAVAL . . . . . Dépt : Mayenne

1ère THESE : "Inéquations variationnelles ; existence, approximation résolution." . . . . . .

. . . . . .

. . . . . .

. . . . . .

2e THESE (Propositions faites par l'Université) : "Fonctions holomorphes dans des espaces de Banach complexes". . . . . .

. . . . . .

. . . . . .

PRESIDENT : M. LIONS

EXAMINATEURS : M. HERVE
M. HAUGAZEAU

Vu et Approuvé

PARIS, le 24 novembre 1972

Le Président

de l'Université de PARIS VI